



DEPARTAMENTO DE BIOQUÍMICA
FACULTAD DE MEDICINA
UNIVERSIDAD AUTÓNOMA DE MADRID

**BÚSQUEDA DE GENES DE SUSCEPTIBILIDAD A
CÁNCER DE MAMA ESPORÁDICO:
ESTUDIOS DE ASOCIACIÓN Y ESTUDIOS FUNCIONALES**

Tesis doctoral que presenta para optar al título de
Doctora por la Universidad Autónoma de Madrid
la Licenciada en Biología y Licenciada en Bioquímica

EVA BARROSO RAMOS

Co-directores de Tesis

Dra. Gloria Ribas Despuig – Investigadora del Grupo de Genética Humana
Dr. Javier Benítez Ortiz – Director del Programa de Genética del Cáncer Humano



Grupo de Genética Humana
Programa de Genética del Cáncer Humano
Centro Nacional de Investigaciones Oncológicas (CNIO)

Los co-directores de tesis, la Dra. Gloria Ribas Despuig, investigadora del Grupo de Genética Humana del Centro Nacional de Investigaciones Oncológicas (CNIO) y el Dr. Javier Benítez Ortiz, director del Programa de Genética del Cáncer Humano del CNIO y profesor honorario del Departamento de Biología Molecular de la Facultad de Ciencias de la Universidad Autónoma de Madrid (UAM),

CERTIFICAN:

Que Eva Barroso Ramos, licenciada en Biología por la Universidad Complutense de Madrid y licenciada en Bioquímica por la Universidad Autónoma de Madrid, ha realizado la presente Tesis Doctoral **“Búsqueda de genes de susceptibilidad a cáncer de mama esporádico: estudios de asociación y estudios funcionales”** y que, a su juicio, reúne plenamente todos los requisitos necesarios para optar al **Grado de Doctora en Biología**, a cuyos efectos será presentada en la Universidad Autónoma de Madrid. El trabajo ha sido realizado bajo su dirección, autorizando su presentación ante el Tribunal Calificador.

Y para que así conste se extiende el presente certificado,

Madrid, abril de 2009.

Vº Bº de los co-Directores de la Tesis:

Dra. Gloria Ribas

Investigadora del Grupo
de Genética Humana
CNIO

Dr. Javier Benítez

Director del Programa de
Genética del Cáncer Humano
CNIO

Vº Bº del Tutor:

Dr. Jesús Cruces

Profesor Titular
Departamento de Bioquímica
Facultad de Medicina, UAM

Esta tesis doctoral ha sido realizada en el Grupo de Genética Humana del Programa de Genética del Cáncer Humano del Centro Nacional de Investigaciones Oncológicas (CNIO), bajo la supervisión de la Dra. Gloria Ribas Despuig y el Dr. Javier Benítez Ortiz.

Las siguientes ayudas han permitido la realización de la presente tesis doctoral:

- Ayuda de Formación de Personal Investigador (FPI) de la Dirección General de Universidades e Investigación de la Comunidad de Madrid, Orden 7580/2003.
- Ayuda de estancia corta de la Dirección General de Universidades e Investigación de la Comunidad de Madrid, Orden 1925/2006.
- Proyecto del Ministerio de Ciencia y tecnología (MCyT09-02), BF12003-03852 (2003-2006).

La frase mas excitante que se puede oír en ciencia, la que anuncia nuevos descubrimientos, no es "¡Eureka!" sino "Es extraño..."

Isaac Asimov

**A mis padres, a mis hermanos
y a Migue**

AGRADECIMIENTOS

Escribir los agradecimientos de la tesis es un ejercicio de recuerdo mayor incluso que la tesis misma, aunque la emoción que suscita se codea con la producida por la escritura de la última frase de las conclusiones, no lo voy a negar. Estas palabras se acompañan, inevitablemente, de cierta nostalgia, pero las sonrisas que provoca el mero recuerdo de los momentos vividos con mis compañeros del CNIO y la ilusión de la nueva etapa que se presenta por delante, me ayudan a iluminar la escritura de este apartado, ya que será sin duda el más leído de la tesis.

Mis primeros agradecimientos, a Gloria y a Javier. Por su comprensión, por su paciencia, por su buen hacer en ciencia, por tantas cosas que han pasado en estos cinco años que nos conocemos. Ahora empieza un nuevo camino, pero su recuerdo y espero que su apoyo, permanecerán siempre. A ellos les agradezco el haber podido realizar la tesis doctoral en su laboratorio. La tesis, un paso en la carrera científica y académica que para muchos supone un trago amargo, o quizá no una situación lo suficientemente agradable como para animarse a continuar en esto, y sin embargo para mí, salvo por los momentos más oscuros por los que todos pasamos, ha sido una experiencia estupenda. Y es que a ellos es a quienes se les debe el ambiente de trabajo tan bueno que he podido disfrutar en este tiempo: el grupo de Genética Humana, el CeGen y el programa de Genética del Cáncer Humano en general, plagado de personas excelentes que han sabido darme la alegría y los consejos que he necesitado y darle al adjetivo de “humano” todo su valor. Glo, qué “mami de ciencia” más estupenda has sido, te echaré mucho de menos, y Javier, qué paciencia has tenido conmigo. Gracias a los dos.

En cuanto a mis compañeros de fatigas, sé por quién empezar, por supuesto, aunque no sé cómo seguir para que nadie se sienta defraudado. Pero me permitirán poner en el lugar que se merece a mi Lari, mi hermana en ciencia, mi despertador personal, mi conciencia, mi serenidad, mi amiga. Espero que sigas en mi vida como has estado hasta ahora, presente cuando te he necesitado, siempre comedida, sensata y alegre, siempre.

A todos los demás, ya saben lo que me gusta un beso y un abrazo, así que no duden que se los daré en cuanto tenga ocasión. Todos me han dado vida en este tiempo, todos tienen un hueco en mi corazón; quizá a partir de ahora y por cómo nos puede alejar la vida, tan sólo en mi recuerdo, pero prometo que será de los mejores. Roger, Charo,

Tais, Dany, Ross, Guille, Iván, Anna, y los que se han ido y que todos recordamos, Emilio y Jesús, todos mis compañeros del CeGen. Qué risas nos hemos echado, qué jaleo a veces, qué paz tantas otras. Espero que al menos echen de menos mis carcajadas, como yo echo de menos a mis “floreillas silvestres” y nuestros momentos de las cuatro de la tarde de los viernes, mis bacterias colonizando la nevera, las dudas estadísticas, la genotipación en grado superlativo, las charlas en la pecera, los achuchones, mis telediarios ya pasados. La otra mitad de mí, Genética Humana, también me ha hecho pasar ratos inolvidables. Gracias a Magda, Ali, Fernando, Ana, Ricardo, Bea, Loren y las nuevas incorporaciones, Maider y Miljana. Me ha encantado compartir con ustedes muchas tardes, muchos congresos, muchas fiestas, muchos *lab meetings*, nunca un problema y sí siempre una sonrisa. Son los mejores. Los demás, compañeros de congresos, *Retreats*, fiestas del CNIO y otras fiestas no oficiales, cómo no mentarlos, vaya risas: Susanna, Alberto, Cristina R., Cristina M., Sergio, Íñigo, Fati, Rocío, Elena, Javi, Laura, Eva, Bárbara..., gracias por haber convertido en geniales tantos ratos. Y a los que han tenido la suerte de no “sufrirme” en las situaciones anteriores, como son Meme, María, Maika, Toya, Miguel, Epidemiología y Citogenética en general, desde aquí les brindo mi mayor cariño. También quiero agradecer la paciencia y el apoyo que mis nuevos compañeros de Endocrinología Genética me han brindado en la consecución de esta tesis.

Finalmente, agradezco a mis padres, Domingo y Blanca, a mis hermanos, Moisés, Alfredo y María, y a mi querido Migue todo el apoyo que me han dado en este tiempo, la confianza de que podía salir algo bueno de todo esto tras el esfuerzo que supone estar lejos de Tenerife. Gracias a mi familia por aguantarme al teléfono cuando lo necesito, gracias por el apoyo económico, gracias por compartir este sueño conmigo pero, por encima de todo, gracias por seguir echándome de menos tras diez años fuera de casa y por no enfadarse porque la vuelta no se presente en un futuro inmediato. A Migue, gracias por formar conmigo una nueva familia y por ilusionarte tanto como yo por esta tesis. En definitiva, gracias a todos por haberme ayudado a soñar.

A todos y cada uno, a mi querido CNIO, un fuerte abrazo.

Eva Barroso Ramos
Madrid, abril de 2009

RESUMEN/ABSTRACT

El cáncer de mama esporádico, que representa hasta el 95% de los casos con cáncer de mama, es una enfermedad compleja causada por cambios comunes en la población en un número indeterminado de genes, en combinación con factores ambientales. La identificación y caracterización de estos cambios comunes han sido el objetivo de la presente tesis doctoral. Se siguió un diseño de estudio de asociación caso-control con una selección de 33 genes candidatos, 19 de ellos implicados en funciones de reparación del ADN, y 14 genes con funciones en el control del ciclo celular, genotipando un total de 169 SNPs en 547 casos de cáncer de mama esporádico y 552 controles de población normal. Los resultados de asociación de genotipo y fenotipo, así como los análisis funcionales realizados han establecido, principalmente, la implicación de dos genes de las familias *FANC* y *E2F* en el desarrollo del cáncer de mama esporádico. Un SNP codificante sinónimo en el gen *FANCD2* se asoció con la enfermedad, y la búsqueda de otras variantes putativas causales identificó un cambio *de novo* en su región promotora que presentó una tendencia a estar asociado con el cáncer de mama. Un SNP en 3'UTR del gen *E2F1*, por su parte, mostró asociación con la enfermedad y se apuntó a un efecto del mismo en el sistema de regulación por microRNAs. Este trabajo, por tanto, ha permitido poner de manifiesto una mayor implicación de los SNPs reguladores en la caracterización de patologías complejas como es el cáncer de mama.

Sporadic breast cancer, which involves up to 95% of all breast cancer patients, has been defined as a complex disease and is caused by common variants in a large and undefined number of genes, together with environmental factors. The identification and characterization of common variants in novel breast cancer genes have been the main objectives of this thesis. A case-control association study was applied using a selection of 33 candidate genes: 19 of them are involved in the DNA repair pathway, and 14 genes in the cell cycle control pathway, genotyping a total of 169 SNPs in a cohort of 547 sporadic breast cancer cases and 552 controls. Results of genotype and phenotype associations, together with functional studies have established the involvement from two genes of the *FANC* and *E2F* families in sporadic breast cancer development. A synonymous coding SNP in *FANCD2* was shown to be associated with the disease, and the screening for other putatively causative variants identified a novel SNP located at the promoter region of this gene, which showed a tendency to be associated with breast cancer. In contrast, a SNP at the 3'UTR of *E2F1* showed an association with disease and evidence pointed out to an effect of this SNP on the microRNA regulation system. Therefore, the obtained results have indicated the important contribution of regulatory SNPs in the characterization of complex diseases, such as breast cancer.

ÍNDICE

CLAVE DE ABREVIATURAS	1
 I. INTRODUCCIÓN	 9
I.1. Características generales del cáncer de mama	11
I.1.1. Descripción del cáncer de mama	11
I.1.1.1. Estructura e histología de la glándula mamaria	11
I.1.1.2. Tipos de cáncer de mama	12
I.1.2. Epidemiología del cáncer de mama y factores de riesgo.....	14
I.2. Genética del cáncer de mama	17
I.2.1. El cáncer de mama familiar y de agregación familiar.....	17
I.2.1.1. Implicación de los genes <i>BRCA1</i> y <i>BRCA2</i> en el cáncer de mama familiar: Estudios de ligamiento	18
I.2.2. El cáncer de mama esporádico.....	20
I.2.2.1. Estudios de asociación en genes candidatos.....	21
I.2.2.2. Estudios de asociación en todo el genoma.....	23
I.2.2.3. Validación de los estudios de asociación.....	24
I.2.2.4. El Consorcio de Estudios de Asociación al Cáncer de Mama.....	25
I.3. Rutas de genes de susceptibilidad a cáncer de mama esporádico.....	26
I.3.1.1. Genes implicados en la reparación del ADN: familia de genes <i>FANC</i> ..	26
I.3.1.2. Genes implicados en el control del ciclo celular: familia de genes <i>E2F</i>	27
 II. OBJETIVOS	 31
II.1. Detección de nuevos genes de susceptibilidad a cáncer de mama esporádico..	33
II.2. Validación de los SNPs que presenten asociación a cáncer de mama	33
 III. MATERIALES Y MÉTODOS	 35
III.1. Material biológico para el estudio	37
III.1.1. Muestras de ADN desde sangre periférica	37
III.1.1.1. Obtención de muestras de sangre periférica y características generales de las muestras	37
III.1.1.2. Extracción de ADN genómico desde sangre periférica.....	40
III.1.1.3. Valoración de la cantidad y calidad del ADN genómico extraído...	40
III.1.1.4. Amplificación del ADN genómico.....	42
III.1.1.5. Cuestionarios y base de datos clínicos	43
III.1.2. Muestras de tejido embebido en parafina.....	44
III.1.2.1. Obtención de muestras de tejido embebido en parafina	44
III.1.2.2. Extracción de ARN total desde tejidos embebidos en parafina	45

III.1.3. Muestras de líneas celulares.....	47
III.1.3.1. Líneas celulares utilizadas	47
III.1.3.2. Extracción de ADN genómico desde líneas celulares.....	47
III.1.3.3. Extracción de ARN total desde líneas celulares	48
III.2. Identificación y selección de genes y polimorfismos.....	50
III.2.1. Selección de genes de interés para el cáncer de mama	50
III.2.2. Bases de datos y selección de SNPs	50
III.2.2.1. Criterios de selección de los SNPs	50
III.2.2.2. Recogida de datos de SNPs de una región genómica	51
III.2.2.3. Definición de bloques de desequilibrio de ligamiento y <i>tag</i> -SNPs	52
III.2.2.4. Búsqueda de SNPs con funcionalidad en el gen	54
III.3. Genotipado de los SNPs seleccionados.....	57
III.3.1. Control de calidad de los ensayos de genotipado	57
III.3.2. PCR	57
III.3.3. <i>TaqMan</i>	59
III.3.3.1. Principios metodológicos de la discriminación alélica por <i>TaqMan</i>	59
III.3.3.2. Ensayo de discriminación alélica por <i>TaqMan</i>	60
III.3.4. <i>Amplifluor</i>	62
III.3.4.1. Principios metodológicos de la discriminación alélica por <i>Amplifluor</i>	62
III.3.4.2. Ensayo de discriminación alélica por <i>Amplifluor</i>	63
III.3.5. <i>dHPLC</i>	65
III.3.5.1. <i>dHPLC</i> como alternativa a <i>TaqMan</i> y <i>Amplifluor</i>	65
III.3.5.2. Principios metodológicos del <i>dHPLC</i>	65
III.3.5.3. Estudio de genotipación a través del <i>dHPLC</i>	66
III.3.6. Secuenciación	67
III.3.7. <i>MassArray iPLEX Gold</i>	68
III.3.7.1. Principios metodológicos de la discriminación alélica por <i>iPLEX</i>	68
III.3.7.2. Ensayo de discriminación alélica por <i>iPLEX</i>	68
III.3.8. <i>Illumina GoldenGate</i>	71
III.3.8.1. Principios metodológicos de la discriminación alélica por <i>Illumina GoldenGate</i>	71
III.3.8.2. Ensayo de discriminación alélica por <i>Illumina GoldenGate</i>	72
III.4. Ensayos funcionales.....	75
III.4.1. Ensayos <i>in silico</i>	75
III.4.2. Ensayos <i>in vitro</i> : clonaje	75
III.4.2.1. Vectores de clonaje.....	75
III.4.2.2. PCR de clonaje	76
III.4.2.3. Digestión del inserto	76
III.4.2.4. Ligación de vector e inserto	76
III.4.2.5. Transformación de células bacterianas competentes	77
III.4.2.6. Extracción del plásmido clonado de bacterias transformadas.....	78
III.4.2.7. Mutagénesis dirigida.....	78

III.4.2.8. Transfección de células eucariotas	79
III.4.2.9. Ensayo de luciferasa	79
III.4.2.10. RT-Q-PCR	80
III.4.3. Ensayos <i>in vivo</i>	82
III.4.3.1. Inmunohistoquímica de piezas de tejido mamario embebido en parafina	82
III.4.3.2. RT-Q-PCR de piezas de tejido mamario embebido en parafina	83
III.5. Análisis estadístico	84
III.5.1. Estudio de asociación del alelo raro	84
III.5.2. Estudio de asociación del genotipo	84
III.5.3. Estudio de asociación del haplotipo y del diplotipo	85
III.5.3.1. Estimación de haplotipos y diplotipos por <i>PHASE</i>	85
III.5.3.2. Análisis estadístico de haplotipos y diplotipos	85
III.5.4. Estudio de asociación del genotipo con variables fenotípicas	86
III.5.5. Estudio estadístico de los ensayos funcionales	86
III.5.5.1. Análisis estadístico de transfección	86
III.5.5.2. Análisis estadístico de inmunohistoquímica y RT-Q-PCR.....	87
 IV. RESULTADOS	 89
IV.1. Generalidades de los datos clínicos.....	91
IV.1.1. Definición de los factores clínicos de confusión.....	91
IV.1.2. Supervivencia en función de las características clínicas.....	92
 IV.2. Estudio de asociación a cáncer de mama esporádico de genes implicados en la reparación del ADN.....	 94
IV.2.1. Genes y SNPs seleccionados de la ruta de reparación del ADN.....	94
IV.2.2. Relación de <i>FANCD2</i> y el cáncer de mama esporádico	95
IV.2.2.1. Estudio de asociación de <i>FANCD2</i>	95
IV.2.2.2. Búsqueda de la variante causal de la asociación de <i>FANCD2</i>	97
IV.2.2.3. Análisis <i>in silico</i> de las variantes putativas causales.....	102
IV.2.3. Estudio de asociación de la familia de genes <i>FANC</i>	103
IV.2.3.1. Estudio de asociación de genotipado de los genes <i>FANC</i>	103
IV.2.3.2. Estudio de asociación de genotipo y fenotipo de los genes <i>FANC</i>	107
IV.2.4. Estudio de otros genes relacionados con la familia de genes <i>FANC</i>	112
IV.2.4.1. Estudio de asociación de genotipado de los genes relacionados con los genes <i>FANC</i>	112
IV.2.4.2. Estudio de asociación de genotipo y fenotipo de los genes relacionados con los genes <i>FANC</i>	114
 IV.3. Estudio de asociación a cáncer de mama esporádico de genes implicados en el control del ciclo celular.....	 118
IV.3.1. Genes y SNPs seleccionados de la ruta de control del ciclo celular.....	118

VI. CONCLUSIONES	155
VII. BIBLIOGRAFÍA	159
VIII. ANEXOS	I
VIII.1. ANEXO I: Información adicional	III
VIII.2. ANEXO II: Publicaciones originadas por la tesis	XX
VIII.3. ANEXO III: Publicaciones relacionadas con la tesis	XXXVIII

CLAVE DE ABREVIATURAS

A continuación se presentan por orden alfabético las abreviaturas utilizadas en la presente tesis, que corresponden a nombres de genes y proteínas, abreviaturas propias de la tesis, o términos científicos descritos inicialmente en lengua inglesa, utilizados comúnmente de esta forma a través de siglas.

ACTB: Actina tipo β (*actin beta*).

AT: Ataxia telangiectasia.

ATM: Proteína mutada en ataxia telangiectasia (*ataxia telangiectasia mutated*).

ATR: Proteína relacionada con ataxia telangiectasia y Rad3 (*ataxia telangiectasia and Rad3 related protein*).

Bcl-2: Proteína de linfomas de células B tipo 2 (*B-cell lymphoma protein 2*).

BCAC: Consorcio de Estudios de Asociación al Cáncer de Mama (*Breast Cancer Association Consortium*).

BER: Reparación por escisión de bases (*base-excision repair*).

BRCA1: Proteína de susceptibilidad a cáncer de mama tipo 1 (*breast-cancer susceptibility gene 1*).

BRCA2: Proteína de susceptibilidad a cáncer de mama tipo 2 (*breast-cancer susceptibility gene 2*).

BRCAX: Proteína de susceptibilidad a cáncer de mama de tipo desconocido (*breast-cancer susceptibility gene X*).

BRG1: Proteína activadora de la transcripción y la división mitótica (*BRM/SWI2-related gene 1*).

BRIP1: Proteína helicasa que interacciona con el extremo C-terminal de BRCA1 (*BRCA1 interacting protein C-terminal helicase 1*). También se denomina BACH1 (*BRCA1-binding helicase-like protein BACH1*).

BUB1B: Proteína de control mitótico (*mitotic checkpoint serine/threonine-protein kinase BUB1 beta*).

CHEK2: Proteína de control del ciclo celular CHK2 (*checkpoint-like protein CHK2*).

CMA: Casos con cáncer de mama de agregación familiar.

CME: Casos con cáncer de mama esporádico.

c-MYC: Proto-oncogén myc (*myc proto-oncogene protein*).

CoCM: Controles de cáncer de mama.

dHPLC: Cromatografía líquida desnaturizante de alto rendimiento (*denaturing high pressured liquid chromatography*).

DL: Desequilibrio de ligamiento (*LD, linkage disequilibrium*).

DNAsa: Enzima que degrada el ADN.

DP1: Compañero de dimerización de E2F tipo 1 (*E2F dimerization partner 1*).

DP2: Compañero de dimerización de E2F tipo 2 (*E2F dimerization partner 2*).

DP4: Compañero de dimerización de E2F tipo 4 (*E2F dimerization partner 4*).

E2F1: Factor de transcripción E2F tipo 1 (*transcription factor E2F1*).

E2F2: Factor de transcripción E2F tipo 2 (*transcription factor E2F2*).

E2F3: Factor de transcripción E2F tipo 3 (*transcription factor E2F3*).

E2F4: Factor de transcripción E2F tipo 4 (*transcription factor E2F4*).

E2F5: Factor de transcripción E2F tipo 5 (*transcription factor E2F5*).

E2F6: Factor de transcripción E2F tipo 6 (*transcription factor E2F6*).

E2F7: Factor de transcripción E2F tipo 7 (*transcription factor E2F7*).

E2F8: Factor de transcripción E2F tipo 8 (*transcription factor E2F8*).

EP300: Proteína p300 asociada a E1A (*E1A-associated protein p300*).

ESE: Región activadora del “*splicing*” (*exonic splicing enhancer*).

ESS: Región silenciadora del “*splicing*” (*exonic splicing silencer*).

FANC: Anemia de Fanconi (*FA, Fanconi anemia*).

FANCA: Proteína de anemia de Fanconi grupo A (*Fanconi anemia group A protein*).

FANCB: Proteína de anemia de Fanconi grupo B (*Fanconi anemia group B protein*).

FANCC: Proteína de anemia de Fanconi grupo C (*Fanconi anemia group C protein*).

FANCD1: Proteína de anemia de Fanconi grupo D1 (*Fanconi anemia group D1 protein*).

FANCD2: Proteína de anemia de Fanconi grupo D2 (*Fanconi anemia group D2 protein*).

FANCE: Proteína de anemia de Fanconi grupo E (*Fanconi anemia group E protein*).

FANCF: Proteína de anemia de Fanconi grupo F (*Fanconi anemia group F protein*).

FANCG: Proteína de anemia de Fanconi grupo G (*Fanconi anemia group G protein*).

FANCI: Proteína de anemia de Fanconi grupo I (*Fanconi anemia group I protein*).

FANCI: Proteína de anemia de Fanconi grupo J (*Fanconi anemia group J protein*).

FANCL: Proteína de anemia de Fanconi grupo L (*Fanconi anemia group L protein*).

FANCM: Proteína de anemia de Fanconi grupo M (*Fanconi anemia group M protein*).

FANCN: Proteína de anemia de Fanconi grupo N (*Fanconi anemia group N protein*).

FMS2: Fragmento de 3'UTR de *E2F1* que porta el sitio 2 de unión de miRNAs mutado en la "secuencia semilla" y rs3213180.

FnoS2: Fragmento de 3'UTR de *E2F1* que no porta el sitio 2 de unión de miRNAs, pero sí rs3213180.

FS2: Fragmento de 3'UTR de *E2F1* que porta el sitio 2 de unión de miRNAs y rs3213180.

GWAS: Estudios de asociación del genoma completo (*genome-wide association studies*).

H2AX: Miembro X de la familia de histonas H2A (*H2A histone family, member X*).

HeLa: Células de carcinoma cervical de la paciente Henrietta Lacks.

HER2/neu: También denominado ErbB-2 o ERBB2. Receptor del factor de crecimiento epidérmico humano tipo 2 (*human epidermal growth factor receptor 2*).

hME: Ensayo MassEXTEND homogéneo (*homogeneous MassEXTEND*).

HWE: Equilibrio de Hardy-Weinberg (*Hardy-Weinberg equilibrium*).

IGF-1: Factor de crecimiento tipo insulina 1 (*insulin-like growth factor 1*).

Ki-67: Proteína identificada con el anticuerpo monoclonal Ki-67 (*antigen identified by monoclonal antibody Ki-67*).

MAF: Frecuencia del alelo menor (*minor allele frequency*).

MALDI-TOF: Desorción-ionización mediante láser asistida por matriz-tiempo de vuelo (*Matrix-Assisted Laser Desorption/Ionization - Time-Of-Flight*).

MCAD: Deshidrogenasa de la coenzima A (*medium-chain acyl-Coenzyme A dehydrogenase*).

MCF-7: Células cancerígenas de mama tipo 7 (*mammary carcinoma cells type 7*).

MDM2: Proteína homóloga a MDM2 de unión a p53 (*p53-binding protein mouse double minute 2 homolog*).

miRNA: MicroRNA.

MRE11: Proteína de recombinación meiótica tipo 11 (*meiotic recombination 11 homolog A*).

MS2: Sitio 2 de unión de miRNAs en *E2F1* mutado.

NBS1: Proteína del síndrome Nijmegen de rotura del ADN (*Nijmegen breakage syndrome 1*).

- OGG1: Glicosilasa del ADN 9-oxoguanina tipo 1 (*8-oxoguanine DNA glycosylase 1*).
- OR: *Odds Ratio*, equivalente a riesgo relativo.
- P53: Proteína supresora de tumor p53 (*p53 tumor supresor protein*).
- PALB2: Compañero y co-localizador de BRCA2 (*partner and co-localizer of BRCA2*).
- PBS: Tampón fosfato salino (*phosphate buffered saline*).
- PcG: Proteína del grupo Polcomb (*polycomb group protein*).
- PCR: Reacción en cadena de la polimerasa (*polymerase chain reaction*).
- PTEN: Proteína homóloga de la fosfatasa y tensina (*phosphatase and tensin homolog*).
- PTEN-MS: Construcción con 3'UTR de *PTEN* con el sitio de unión de miRNAs mutado en la "secuencia semilla".
- PTEN-WS: Construcción con 3'UTR de *PTEN* con el sitio de unión de miRNAs intacto.
- RA: Receptor de andrógenos.
- RAD50: Proteína de recombinación A tipo 50 (*recombination protein A 50*).
- RAD51: Proteína de recombinación A tipo 51 homóloga a RecA de *E. coli* (*recombination protein A 51, RecA homolog E. coli*).
- RB1: Proteína asociada al retinoblastoma tipo 1 (*retinoblastoma-associated protein 1*).
- RBL1: Proteína parecida a RB tipo 1, también denominada p107 (*retinoblastoma-like protein 1*).
- RBL2: Proteína parecida a RB tipo 2, también denominada p130 (*retinoblastoma-like protein 2*).
- RE: Receptor de estrógeno.
- RISC: Complejo de silenciamiento inducido por ARN, que forma una estructura con los microRNAs (*RNA-induced silencing complex*).
- RNAsa: Enzima que degrada el ARN.
- RP: Receptor de progesterona.
- RT-Q-PCR: PCR a tiempo real tras una transcripción reversa (*reverse transcriptase quantitative PCR*).
- SAP: Fosfatasa alcalina de gamba (*shrimp alkaline phosphatase*).
- S1: Sitio 1 de unión de miRNAs en *E2F1*.
- S2: Sitio 2 de unión de miRNAs en *E2F1*.
- SNP: Polimorfismo por cambio de un único nucleótido (*single-nucleotide polymorphism*).

SNPr: SNP regulador.

STK6: Quinasa de serina y treonina tipo 6 (*serine/threonine kinase 6*, Aurora-A).

Tag-SNPs: SNPs que sirven de marcadores genéticos.

TNM: Sistema de estadificación del tumor, según T, tamaño del tumor, N, metástasis a ganglios linfáticos axilares, y M, metástasis a tejidos distantes de la mama.

T_m: Temperatura de “*melting*” (fusión) o disociación del ADNdc.

USP1: Peptidasa ubiquitín-específica tipo 1 (*ubiquitin specific peptidase 1*).

UTR: Región no traducida de ARN a proteína (*untranslated region*).

VDR: Receptor de vitamina D (*vitamin D receptor*).

I. INTRODUCCIÓN

I.1. CARACTERÍSTICAS GENERALES DEL CÁNCER DE MAMA

I.1.1. Descripción del cáncer de mama

I.1.1.1. Estructura e histología de la glándula mamaria

La glándula mamaria está formada por tres tipos de tejidos: glandular de tipo túbulo-alveolar, conjuntivo que conecta los lóbulos, y adiposo que ocupa los espacios interlobulares. Cada glándula mamaria se compone de quince a veinte lóbulos que desembocan, a través de un sistema tubular, en los conductos galactóforos que vierten leche en el pezón tras el parto. Los lóbulos mamarios están constituidos por numerosos lobulillos que se encuentran unidos entre sí por tejido conjuntivo, vasos sanguíneos, vasos linfáticos y por su sistema excretor, los conductos galactóforos. Los lobulillos, a su vez, están formados por diez a cien acinos, cada uno con un conducto excretor denominado conducto terminal. Los acinos están estructurados por un conjunto de células secretoras que producen la secreción láctea y conforman una cavidad a la cual vierten dicha secreción. Éstos se encuentran rodeados de células mioepiteliales y capilares sanguíneos que participan activamente en el proceso de secreción y eyección de la leche (*Figura 1*).

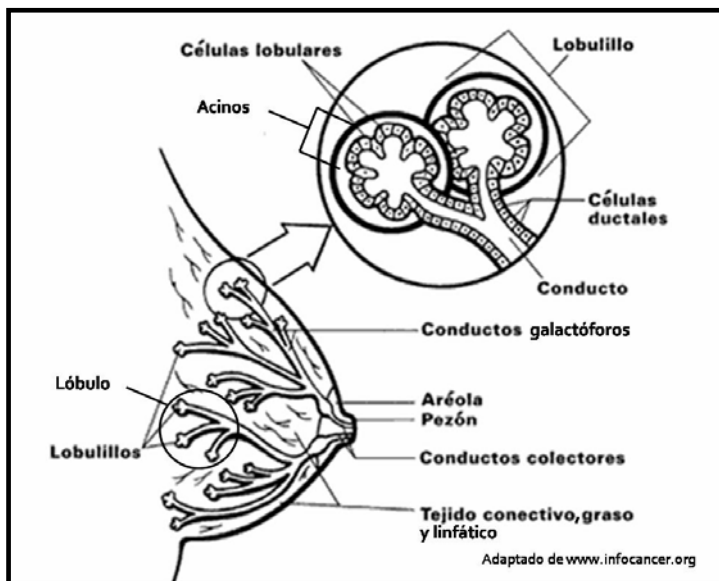


Figura 1: Estructura del tejido mamario: lóbulo, lobulillos, acinos, conductos galactóforos, conductos colectores y pezón, rodeados de tejido conectivo, graso y linfático.

Los lobulillos están formados por células columnares basófilas y miofibrillas con una alta sensibilidad a la oxitocina, y se encuentran inmersos en tejido conectivo sin

fibras elásticas. Los conductos, por su parte, están constituidos por un epitelio cilíndrico o cúbico envuelto por células mioepiteliales, rodeado de tejido conectivo y fibras elásticas. En ambos casos las células epiteliales secretoras constituyen la capa luminal, mientras que las células mioepiteliales representan la capa basal.

La estructura de la glándula mamaria varía con la edad y está influenciada por los ciclos menstruales, el embarazo y la lactancia, estadio en el que adquiere su mayor desarrollo, atrofiándose tras la menopausia y la ausencia de los estímulos hormonales de estrógeno y progesterona (Cotran *et al.*, 2000).

I.1.1.2. Tipos de cáncer de mama

La mayoría de los tumores que se producen en la mama son benignos, debidos a formaciones fibroquísticas que pueden resultar dolorosas pero no cancerígenas. Sin embargo, en los casos en los que se produce un crecimiento anormal y desordenado de las células mamarias, se origina un tumor maligno en el tejido mamario, o cáncer de mama. El adenocarcinoma representa el tipo más común de cáncer de mama invasivo, deriva del epitelio parenquimal mamario y se caracteriza por su capacidad de invasión de tejidos adyacentes y por una marcada tendencia a metastatizar a tejidos distantes tales como hueso, pulmón, pleura, hígado, glándulas adrenales, ovarios, piel y cerebro.

El cáncer de mama se clasifica de diversas formas: en función de su carácter familiar o de aparición esporádica, del lugar de la mama donde se produce el crecimiento anormal de las células, del tipo de célula afectada, del estadio del cáncer y de las características moleculares del tumor.

Los tipos de cáncer de mama, acorde a su componente genético familiar o esporádico, serán comentados más adelante en el apartado I.2.

Los dos tipos de cáncer de mama descritos según el tipo de célula malignizada son: carcinoma ductal, donde están afectadas las células de los ductos y supone el 90% de los casos de cáncer de mama, y carcinoma lobulillar, donde se malignizan las células de los acinos glandulares o células lobulillares y representa el 10% restante de los casos.

Atendiendo al lugar en el que se encuentra el carcinoma, se pueden diferenciar dos tipos de cáncer de mama: carcinoma *in situ*, que permanece confinado a la luz de los ductos o de los acinos, y carcinoma invasivo o infiltrante, en el que las células proliferan

lo suficiente como para romper la membrana basal y extenderse infiltrando los tejidos que rodean a los ductos y acinos, como son los tejidos conectivo y adiposo, y vasos sanguíneos y linfáticos (Cotran *et al.*, 2000).

Según el estadio del cáncer y el sistema de estadificación TNM creado por la AJCC (American Joint Committee on Cancer) en colaboración con la UICC (Internacional Union Against Cancer) (Greene and Sobin, 2009), se establecen índices numéricos que indican la extensión progresiva de la enfermedad: T, que indica la extensión del tumor, N, que establece la infiltración del tumor a los ganglios linfáticos axilares, adyacentes a la mama, y M, que indica la presencia de metástasis a otros tejidos distantes de la mama (detalles en el *Anexo I.1.*). Además, atendiendo al grado de diferenciación del tumor, se han creado las siguientes categorías: bien diferenciado (grado I) de carácter más leve y con mejor pronóstico, moderadamente diferenciado (grado II) y pobremente diferenciado (grado III) que corresponde al tipo más agresivo. Ambos elementos, el valor TNM y el grado de diferenciación, han sido reconocidos como potentes factores pronósticos de la enfermedad (Oldenburg *et al.*, 2007).

Una clasificación que se está empezando a aplicar en los últimos años sigue un criterio genético del tumor, acorde al patrón de expresión génica diferencial observada en un estudio que incluía cerca de 25.000 genes de expresión constitutiva en el tejido mamario en un total de 261 tumores de cáncer de mama esporádico. El análisis de agrupamiento jerárquico de los tumores según la expresión génica ha evidenciado la heterogeneidad del cáncer de mama, que se ha subdividido molecularmente en cinco tipos diferentes: tumor basal, tumor con sobre-expresión de ERBB2, tumor luminal tipo A, tumor luminal tipo B y tumor con expresión similar al tejido mamario normal (Sorlie *et al.*, 2001; Sorlie *et al.*, 2003). Estos subtipos de tumor mamario presentan diferencias clínicas evidentes. Así, se ha observado que la mayoría de los tumores de pacientes de cáncer familiar con mutación en *BRCA1* pertenecen al grupo de tumores basales, presentando una pérdida de expresión de receptor de estrógenos (RE), de receptor de progesterona (PR), una falta total de amplificación de *ERBB2* y un peor pronóstico (Grushko *et al.*, 2002; van 't Veer *et al.*, 2002). Estos tumores se han denominado comúnmente tumores “triple-negativos” (Irvin and Carey, 2008). Los tumores del subtipo luminal A, por su parte, están asociados a un mejor pronóstico, mientras que los

luminales B muestran un pronóstico intermedio (Sorlie *et al.*, 2003). Los tumores *BRCA1* y *BRCA2* negativos (*BRCAX*) se caracterizan, por su parte, en grupos divergentes acorde a la expresión de un grupo numeroso de genes, lo cual indica la heterogeneidad histopatológica de estos tumores, reforzando la idea de que múltiples eventos genéticos intervienen en su origen (Hedenfalk *et al.*, 2003).

I.1.2. Epidemiología del cáncer de mama y factores de riesgo

El cáncer de mama es el cáncer más frecuente entre las mujeres de países desarrollados, pues supone el 23% de todos los cánceres (Parkin *et al.*, 2005), y representa la localización más importante de cáncer en mujeres entre 20 y 79 años. En hombres, el cáncer de mama también puede estar presente, aunque tan sólo representa un 1% de todos los cánceres de mama diagnosticados. El cáncer de mama muestra una prevalencia¹ del 17,9%, seguido del cáncer colorectal (11,5%) y del cáncer de próstata (9,6%) y, si bien su incidencia² (1,15 millones en el año 2002) se sitúa detrás del cáncer de pulmón (1,35 millones) al considerar ambos sexos, ésta crece anualmente un 0,5%, lo cual justifica la importancia de su estudio en términos de salud pública. La investigación en cáncer de mama, el desarrollo de campañas de control del mismo y un mejor tratamiento está potenciando la alta supervivencia que se observa: un 89% a los cinco años, con un promedio de 75% de supervivencia total en países desarrollados y un 57% en países en vías de desarrollo. Sin embargo, a pesar de estos avances, el cáncer de mama sigue siendo la causa más frecuente de muerte por cáncer entre mujeres, llegando hasta un 14% de fallecimientos (Parkin *et al.*, 2005).

En España, por su parte, las tasas de incidencia y mortalidad del cáncer de mama distan de los datos descritos anteriormente, al ser una de las más bajas de los países desarrollados. Datos recogidos por la Asociación Española Contra el Cáncer (AECC) indican que en España se diagnosticaron 16.000 nuevos casos de cáncer de mama en el año 1.998, que produjeron la muerte de 6.000 mujeres, estableciendo una incidencia de 67 casos por 100.000 habitantes, la menor de la Unión Europea (Lopez-Abente *et al.*,

¹ Prevalencia: estimación del número de casos de cáncer vivos en un determinado momento.

² Incidencia: número de casos nuevos de una enfermedad en una población y período determinados, expresado en número de casos por año o como una tasa de afectados por cada 100.000 habitantes por año.

2000) (Figura 2). Datos actualizados a fecha de 2.002 cifraron en 50,9 casos/100.000 habitantes/año la incidencia de cáncer en España, la mayoría diagnosticados entre los 35 y 80 años, estableciendo un máximo de casos entre los 45 y 65 años.

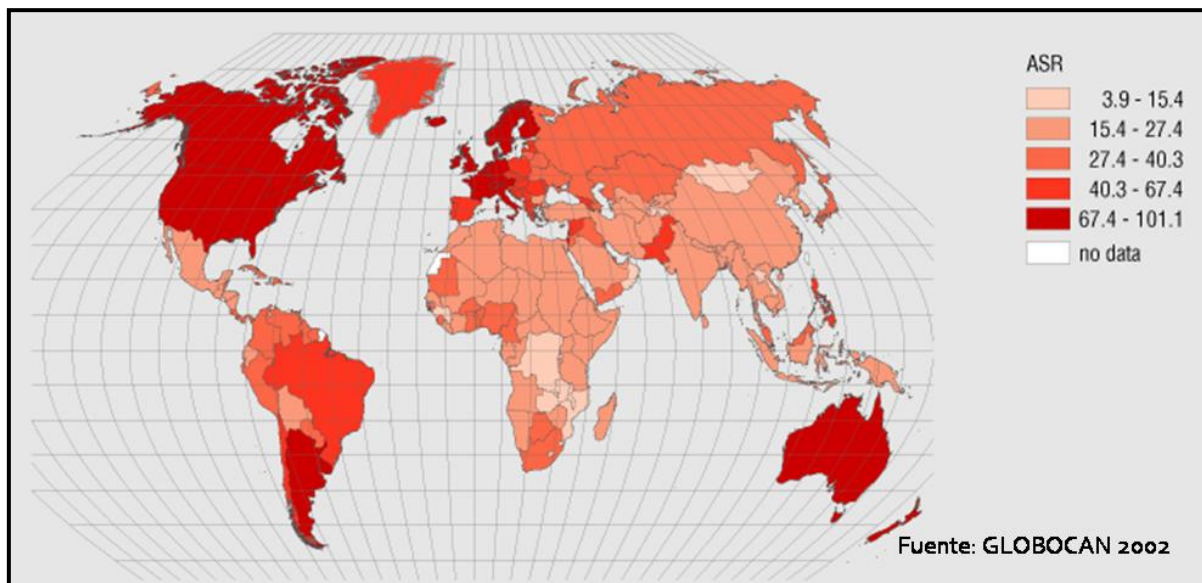


Figura 2: Incidencia del cáncer de mama en el mundo en el año 2.002. España está señalada en color naranja intenso y se sitúa en una incidencia entre 40,3 y 67,4 casos por 100.000 habitantes en 2.002.

El grado de riesgo a padecer cáncer de mama no es homogéneo entre la población general, aunque únicamente el 12% de los pacientes con cáncer de mama presenta un factor de riesgo identificable. Los factores relacionados con el cáncer de mama son los siguientes (Cuzick, 2008; Oldenburg *et al.*, 2007):

1. Contribución genética: tener una historia familiar de cáncer de mama o ser portador de cambios en genes que aportan susceptibilidad a cáncer de mama.
2. Factores demográficos: habitar en países occidentales, ser mujer, envejecer y pertenecer a un nivel socio-económico bajo.
3. Factores endógenos: haber tenido una menarquia temprana (<12 años), llegar a la menopausia a una edad elevada (>54 años), no tener hijos o tener el primer hijo a una edad elevada, no dar el pecho y tener una baja actividad física.
4. Factores exógenos: utilizar anticonceptivos orales, tratarse con terapia hormonal sustitutiva o haber sido expuesto a radiación ionizante antes de la adolescencia.
5. Características físicas: tener obesidad tras la menopausia, tener niveles elevados de IGF-1, tener una historia atípica de tumores benignos en la mama, presentar

un tejido mamario denso en las mamografías y tener una alta densidad ósea en las mujeres postmenopáusicas.

6. Factores dietéticos: tomar alcohol, bajas dosis de folato y altas dosis de grasas insaturadas y carnes rojas.

Estudios realizados en inmigrantes provenientes de países en desarrollo que se han trasladado a países desarrollados han evidenciado el incremento de la incidencia de cáncer de mama en esta población (Kliewer and Smith, 1995), lo que indica claramente que los factores ambientales anteriormente citados tienen una contribución importante en el riesgo a padecer cáncer de mama. De igual modo, aquellos factores contrapuestos a los arriba indicados van a ser considerados factores de protección frente al cáncer de mama y son considerados en las campañas de prevención.

La mayoría de los factores anteriores, tanto de riesgo como de protección, están relacionados con los niveles de las hormonas sexuales y el tiempo de exposición a los que está sometida la glándula mamaria, siendo el estrógeno el factor considerado como mayor modificador del riesgo a padecer cáncer de mama (Mitrinen and Hirvonen, 2003). Se han propuesto dos mecanismos mediante los cuales el estrógeno está implicado en el cáncer de mama: el primero establece que la respuesta que provoca el estrógeno en su tejido diana es la proliferación celular, lo cual incrementa las posibilidades de daño genético y carcinogénesis (Russo *et al.*, 2000); el segundo propone el potencial genotóxico de los metabolitos del estrógeno, como es el estrógeno hidroxilado o catecol-estrógeno (Cavalieri *et al.*, 2000).

Por otra parte, tener una historia familiar de cáncer de mama se ha establecido como un importante factor de riesgo. Así, por cada familiar de primer grado de parentesco con cáncer de mama, se eleva el riesgo de padecer la enfermedad al doble, condicionado además por la edad temprana de aparición del cáncer, la presencia de cáncer de mama bilateral o de una historia clínica de tumores benignos en la mama (Thompson and Easton, 2004).

I.2. GENÉTICA DEL CÁNCER DE MAMA

Como ocurre con otros cánceres, el cáncer de mama se ha considerado una enfermedad genética en la que mutaciones en genes reparadores del ADN, controladores de la entrada de apoptosis, reguladores del ciclo celular y otras funciones de importancia en el control de la proliferación celular, llevan al proceso de malignización celular que desemboca en el desarrollo del cáncer (Vogelstein and Kinzler, 2004). Las células cancerígenas llegan, a través de los eventos de mutación anteriores, a un fenotipo con seis características que las diferencian de las células normales: ignoran las señales de parada de proliferación, ignoran las señales de diferenciación, presentan una capacidad de proliferación indefinida, evaden la apoptosis, tienen capacidad de invasión y tienen capacidad de angiogénesis (Ponder, 2001).

Asimismo, el cáncer de mama se considera una enfermedad compleja y heterogénea, debida tanto a un componente genético como ambiental. El componente genético está representado por un número indeterminado de cambios en el genoma que interactúan entre ellos y con el ambiente y que llevan a un riesgo estimado de padecer la enfermedad según un modelo poligénico (Pharoah *et al.*, 2002). El componente ambiental, por su parte, está constituido por los factores de riesgo y protección que se señalan en el apartado I.1.2., y que pueden conducir a padecer o no la enfermedad. El cáncer de mama, por tanto, varía enormemente en su comportamiento clínico, la apariencia morfológica y las alteraciones moleculares que presenta, de modo que diferentes tipos de cáncer de mama van a tener perfiles de riesgo característicos, lo que puede evidenciar que siguen una etiología diferente. La descripción del componente genético y ambiental del desarrollo del cáncer de mama se ha convertido en un objetivo fundamental en el estudio del mismo (García-Closas *et al.*, 2008).

I.2.1. El cáncer de mama familiar y de agregación familiar

El cáncer de mama se ha descrito en ocasiones como una enfermedad hereditaria en la que un factor de gran importancia en el riesgo a padecer cáncer de

mama es la historia familiar que se tenga para esta enfermedad. Ello indica un alto componente genético heredado que incrementa el riesgo a padecer cáncer de mama con una alta probabilidad, en lo que se ha venido a denominar cáncer de mama familiar, y que sigue un patrón de herencia mendeliana (Bradbury and Olopade, 2007).

La clasificación de cáncer de mama familiar se ha indicado para aquel que afecta a tres o más familiares de primer grado de parentesco, con edad de diagnóstico temprana y presencia de cáncer de ovario en algunos casos, características que identifican a las familias de alto riesgo (Osorio *et al.*, 2000). Por otro lado, se considera cáncer de mama de agregación familiar al que afecta a dos miembros de la misma familia con una relación de primer grado de parentesco, con una edad de diagnóstico más avanzada y ausencia de casos con cáncer de ovario. Estas familias suponen el 5-10% de todos los cánceres de mama diagnosticados.

I.2.1.1. Implicación de los genes *BRCA1* y *BRCA2* en el cáncer de mama familiar: Estudios de ligamiento

Estudios realizados utilizando familias de alto riesgo para cáncer de mama han llevado a la detección de mutaciones germinales en genes de susceptibilidad a cáncer de mama de alta o media penetrancia³. Éstos pertenecen, fundamentalmente, al grupo de genes supresores de tumor o “*care-takers*”, que actúan como sensores de daño en el ADN y participan en el proceso de reparación del mismo. Los genes más importantes relacionados con el cáncer de mama son *BRCA1*, localizado en el cromosoma 17q21 (Miki *et al.*, 1994) y *BRCA2*, localizado en 13q12 (Wooster *et al.*, 1995), éste último también identificado como *FANCD1* (Howlett *et al.*, 2002; Stewart and Elledge, 2002) (Antoniou *et al.*, 2003). Ambos genes han sido descritos como la mayor causa de agregación familiar a cáncer de mama a través de estudios de ligamiento (Cox *et al.*, 2007; Rahman *et al.*, 2007; Renwick *et al.*, 2006; Seal *et al.*, 2006; Walsh *et al.*, 2006). Los estudios de ligamiento basados en familias con un alto número de afectados para una enfermedad han sido de gran utilidad para la identificación de genes que intervienen en desórdenes monogénicos, como ocurre con el cáncer de mama familiar. La aplicación de estos estudios de ligamiento requiere de la utilización de marcadores genéticos con el

³ Penetrancia: Es la proporción de individuos con un genotipo específico que se manifiesta a nivel de fenotipo.

fin de identificar regiones que segregan con la enfermedad en las familias estudiadas y que, por tanto, contienen la variante genética causal de la misma (Hirschhorn and Daly, 2005). *BRCA1* fue el primer *locus*⁴ encontrado que presentaba segregación con el cáncer de mama en un grupo de familias con múltiples casos y una edad temprana de fallecimiento (Hall *et al.*, 1990). La selección de familias con casos de cáncer de mama en hombres llevó, años más tarde, a la detección del *locus* de *BRCA2/FANCD1* (Wooster *et al.*, 1994) (Tavtigian *et al.*, 1996). A nivel germinal, *BRCA1* (Miki *et al.*, 1994) y *BRCA2/FANCD1* aparecen mutados en el 20-35% de los casos de cáncer de mama hereditario y representan menos del 5% total de casos de cáncer de mama diagnosticados, estableciendo una alta susceptibilidad a padecer la enfermedad a lo largo de la vida, en combinación con un menor riesgo para otros tipos de cáncer (Antoniou *et al.*, 2003). El riesgo acumulado para padecer cáncer de mama se ha estimado en un 46% para portadoras de mutación en *BRCA1*, mientras que portadoras de mutación en *BRCA2/FANCD1* presentan un riesgo del 43% (Chen *et al.*, 2006). Sin embargo, los riesgos estimados indicados se ven condicionados por la localización de las mutaciones en el gen (Thompson *et al.*, 2002), por el efecto de genes modificadores de riesgo (Antoniou *et al.*, 2007), y por el tipo de población. Así, por ejemplo, en población española estos valores se han estimado en un 52% para portadoras de mutación en *BRCA1* y en un 47% para portadoras de mutación en *BRCA2/FANCD1*, a la edad de 70 años (Milne *et al.*, 2008).

Sin embargo, tras el largo tiempo transcurrido desde estos descubrimientos, no se han podido localizar otros genes asociados con la predisposición a cáncer de mama hereditario en familias que no tienen mutaciones en *BRCA1* o *BRCA2/FANCD1*, las denominadas *BRCAX*, utilizando los estudios de ligamiento. Ello está claramente relacionado con la alta heterogeneidad de la enfermedad, en la que múltiples *loci* de alto o medio riesgo intervienen en el mismo fenotipo para el cáncer de mama familiar. Esto podría ser solventado tras una subclasificación de las familias con el fin de establecer grupos de muestras más homogéneas (Rosa-Rosa *et al.*, 2009).

Las proteínas *BRCA1* y *BRCA2/FANCD1* están implicadas en múltiples rutas celulares. *BRCA1* se ha visto relacionada con el control del ciclo celular, interviene en el

⁴ *Locus*: En plural *loci*, hace referencia a cualquier sitio polimórfico del genoma.

remodelaje de la cromatina en eventos de activación de la expresión génica, y tiene funciones de ubiquitinación y, por tanto, activación de otras proteínas. BRCA2/FANCD1 por su parte, es menos conocida, y la función principal que desempeña es la que comparte con BRCA1, participando en la reparación de daño en el ADN por la rotura de la doble hebra a través del mecanismo de recombinación homóloga (Moynahan *et al.*, 1999). Ambas proteínas, BRCA1 y BRCA2/FANCD1, interactúan con la familia de proteínas FANC y con RAD51, un componente clave en el mecanismo de reparación del ADN por recombinación homóloga, con el fin de formar un complejo nuclear posterior al evento de daño del ADN que permitirá la reparación del mismo (Figura 3) (Davies *et al.*, 2001; Marmorstein *et al.*, 1998; Moynahan *et al.*, 2001; Yoshida and Miki, 2004; Yuan *et al.*, 1999). Esto supone que existe una relación entre el mantenimiento de la integridad genómica, la actividad de supresión de tumor y la predisposición a padecer cáncer de mama hereditario (Jasin, 2002; Scully and Livingston, 2000) al haberse evidenciado que la pérdida de actividad de BRCA1 y BRCA2 reduce la eficiencia del mecanismo de reparación del ADN por recombinación homóloga (Moynahan *et al.*, 2001).

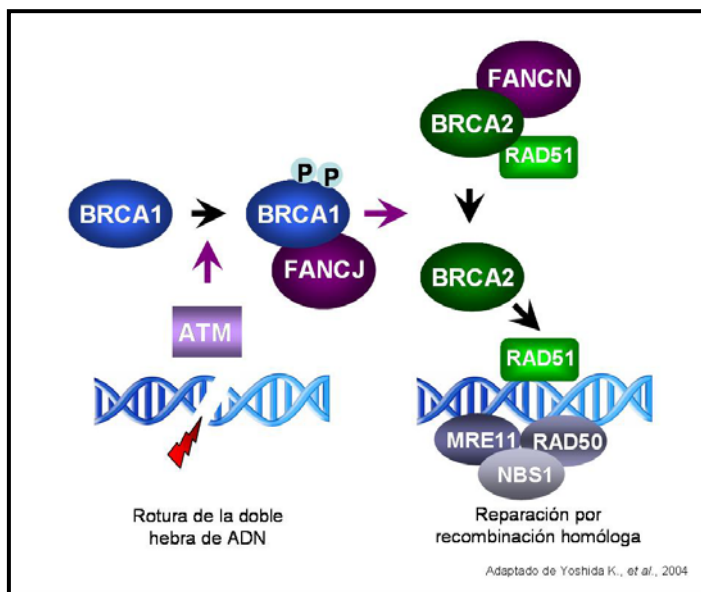


Figura 3: Descripción de las funciones de BRCA1 y BRCA2/FANCD1 en la reparación del ADN por recombinación homóloga. BRCA1 se fosforila por ATM tras la detección del daño en la doble hebra del ADN y se forma un complejo con FANCD1; BRCA1 libera a RAD51 de BRCA2-FANCD1 y activa la cascada de reparación; RAD51 se traslada al complejo nuclear de reparación junto con BRCA2/FANCD1 y BRCA1, donde realizan la reparación por recombinación homóloga en combinación con el complejo RMN (RAD50, MRE11 y NBS1).

I.2.2. El cáncer de mama esporádico

El 95% de los casos de cáncer de mama pertenecen al grupo que se ha venido a denominar cáncer de mama esporádico, para el que no existe una historia familiar de

cáncer evidente, como la indicada en el apartado I.2.1. Éste es considerado un desorden poligénico de alta heterogeneidad (Carter, 2001; Comings *et al.*, 2003), para el que existen genes de bajo riesgo o baja penetrancia, en los que hay presentes variantes comunes en la población general que van a producir una susceptibilidad leve a la enfermedad, en combinación con factores de riesgo ambientales (Antoniou and Easton, 2003; Antoniou *et al.*, 2001; Comings *et al.*, 2003; Risch, 2001). Sin embargo, todavía no está claro cuántos de estos alelos existen para el cáncer de mama y qué proporción de predisposición genética a la enfermedad se atribuye a los mismos. Para éstos, se asumen riesgos relativos menores de 1,5, para lo cual deben participar cientos de variantes comunes (con frecuencias >5%) o miles de variantes raras, a fin de producir la enfermedad (Pharoah *et al.*, 2004). Así pues, en un estudio de mujeres que portan 14 alelos de riesgo, se ha establecido el riesgo de padecer cáncer de mama, aproximadamente, en seis veces más, comparadas con mujeres que no los llevan, indicando que, si bien los riesgos individuales son pequeños, los alelos de riesgo actúan según un modelo multiplicativo (Pharoah *et al.*, 2008). De igual modo, deben existir variantes que protegen frente a la enfermedad y que deben ser tomadas en cuenta en el estudio de riesgo global.

Este modelo no explica el elevado riesgo de padecer cáncer de mama en familias con ausencia de mutación en los genes *BRCA1* y *BRCA2/FANCD1*, pero sí se puede aplicar al estudio de una proporción importante de la incidencia total de cáncer de mama. Por tanto, desde el punto de vista de salud pública, la búsqueda y detección de los genes y las variantes genéticas de baja susceptibilidad que intervienen en el componente genético del cáncer de mama esporádico es de gran importancia, puesto que puede ayudar a definir programas preventivos para el cáncer de mama en mujeres portadoras de alelos de riesgo, aunque su aplicación práctica en la actualidad es todavía cuestionable (Pharoah *et al.*, 2008).

I.2.2.1. Estudios de asociación en genes candidatos

Todos los estudios de asociación tienen como elemento común la genotipación de variantes genéticas comunes en una población con un fenotipo definido como puede ser el cáncer de mama. En su forma más simple, se establece la comparación de las

frecuencias de alelos o genotipos de una variante en particular entre casos con la enfermedad y controles sanos. El fin de este análisis, por tanto, es la detección de una correlación entre el genotipo y el fenotipo, estableciendo una relación entre la variante genética estudiada y la enfermedad.

Los polimorfismos de cambio de un único nucleótido (SNPs, *single-nucleotide polymorphisms*) son considerados la variación más común en la secuencia genómica, donde se produce la sustitución de una base por otra de forma estable. Debido a su frecuente presencia en el genoma, su carácter bialélico en muchos casos y su fácil detección por el gran desarrollo de las tecnologías de genotipado de los últimos años, los SNPs se han destacado como el elemento fundamental de análisis en los estudios de asociación caso-control. Por definición, un SNP tiene una frecuencia del alelo menor (MAF, *minor allele frequency*) mayor del 1% en, al menos, una población (Risch, 2000) y no altera la función o la expresión del gen de manera notable. De este modo, se reserva el concepto de “mutación” para aquellas variantes raras, con una frecuencia menor al 1% indicado. Éstas muestran una particular penetrancia asociada a un fenotipo sintomático de una enfermedad determinada que responde a un desorden genético monogénico clásico, como son las mutaciones identificadas en *BRCA1* y *BRCA2/FANCD1*. Asimismo, los SNPs de un mismo cromosoma no se heredan al azar, sino en una estructura de combinación de alelos, formando un bloque de haplotipo. El genoma, pues, se organiza en bloques de desequilibrio de ligamiento (DL)⁵ (LD, *linkage disequilibrium*), entre los cuales hay regiones de rotura de bloques de DL (Bonnen *et al.*, 2002). Así, la mayoría de las variantes genéticas comunes pueden ser evaluadas en estudios de asociación usando tan sólo unos pocos cientos de miles de SNPs o *tag*-SNPs, aquellos SNPs que se utilizan como marcadores moleculares, que están correlacionados con variantes cercanas con las que comparten su misma frecuencia alélica y que se heredan en su mismo bloque de DL (Hinds *et al.*, 2005).

Acorde a lo expuesto, los estudios de asociación se han basado en la hipótesis de “variante común – enfermedad común”, según el modelo poligénico aplicado para la enfermedad compleja que es el cáncer de mama.

⁵ DL: Desequilibrio de ligamiento es la asociación no azarosa de alelos de diferentes polimorfismos que están relacionados en una población.

En los estudios de asociación en genes candidatos, se realiza una selección de genes acorde a su interés biológico y a su posible relación con la patología (Hirschhorn and Daly, 2005; Oldenburg *et al.*, 2007). Los criterios de selección de SNPs en los estudios de asociación caso-control siguen dos tipos estudios no excluyentes:

1. Análisis directo: selección de SNPs con una posible funcionalidad en la expresión génica o que están conservados filogenéticamente.
2. Análisis indirecto: como se ha indicado, se establece la selección de SNPs como marcadores genéticos, o *tag*-SNPs, que se encuentran en distintos bloques de DL dentro de un gen y que, por tanto, permitirán recoger la mayor parte de la variabilidad genética de esta zona.

I.2.2.2. Estudios de asociación en todo el genoma

En el caso de que los defectos fisiológicos de una enfermedad compleja sean desconocidos o al menos no se hayan definido en su totalidad, como es el caso del cáncer de mama, la utilización de estudios de asociación caso-control a través de genes candidatos no permitirá definir la totalidad de la base genética de la enfermedad. En estos casos, los estudios de asociación en todo el genoma (GWAS, *genome-wide association studies*) se han destacado como una herramienta de gran utilidad. Éstos se definen como un estudio de asociación que interroga a la mayor parte del genoma con el fin de encontrar la variante genética causal de la enfermedad, pero sin establecer una presunción previa de la localización de la misma. Para ello, se realiza el estudio de miles de SNPs, lo cual supone una situación ideal de detección de variantes de riesgo para el cáncer de mama dado que se evita el sesgo que supone elegir previamente las variantes que se estudian (Hirschhorn and Daly, 2005). Sin embargo, el número de muestras requeridas en estos estudios se ve determinada por la MAF y el riesgo de la enfermedad que se quiera detectar. Así, para riesgos relativos del orden de 1,5, con una MAF en un rango de 10-40%, haría falta una población de 1.000 casos y 1.000 controles para poder detectar una asociación estadísticamente significativa (a un nivel de 10^{-4}), con un poder estadístico del 90%. La inclusión del concepto de *tag*-SNPs ha simplificado la realización de este tipo de estudios al aplicarse el estudio de asociación por aproximación indirecta. Se considera que los *tag*-SNPs con una correlación de $r^2 > 0,8$ (ver apartado III.2.2.3 para

más detalle) son suficientes para una cobertura adecuada de la variabilidad genética de una región determinada (Wang *et al.*, 2005). Se ha estimado, asimismo, que la genotipación de 200.000-500.000 *tag*-SNPs es suficiente para identificar aquellos SNPs con una MAF mayor o igual al 5% (Hinds *et al.*, 2005). Así, cualquier variante común en el gen que incrementa el riesgo para la enfermedad será detectada a través del incremento del haplotipo que porta esta variante en el grupo de casos, comparado con los controles sanos. Este procedimiento se ha aplicado en el estudio de susceptibilidad a cáncer de mama esporádico en una serie de trabajos que han reportado la identificación de nuevos genes y regiones intergénicas asociadas a la patología y que no hubieran sido descubiertas de otro modo (Easton *et al.*, 2007; Pharoah *et al.*, 2007).

I.2.2.3. Validación de los estudios de asociación

La utilización de un número reducido de muestras y de SNPs en los estudios de asociación limita el poder de detección de *loci* con un efecto real en la patología estudiada y se detectan asociaciones que resultan en falsos positivos (error tipo I), no confirmados en trabajos posteriores (Consortium, 2006). Sin embargo, la ausencia de replicación de resultados de asociación puede ser debida también a un poder estadístico inadecuado para este estudio, generando falsos negativos (error tipo II). La publicación indiscriminada de resultados de asociación en enfermedades complejas que no se han reproducido en posteriores análisis, así como la limitada publicación de resultados negativos, ha creado un cierto desprestigio en el campo de los estudios de asociación (Dunning *et al.*, 1999). Con el fin de solventar estos inconvenientes, se han desarrollado diversas estrategias que permiten la validación de los resultados obtenidos. Éstas son las siguientes (McCarthy *et al.*, 2008):

1. Ampliación del tamaño muestral de la misma población, con el fin de verificar si se mantiene la asociación detectada.
2. Replicación de la asociación en una población independiente y con una tecnología de genotipado diferente, si es posible. Esta herramienta permite definir asociaciones reproducibles que se consideran entonces verdaderas.
3. Aplicación de estudios de asociación en múltiples fases, en los que aquellos SNPs asociados para la patología en la primera fase con un grupo de muestras

con un fenotipo muy definido, son utilizados para una segunda fase con un grupo diferente de muestras con un fenotipo más complejo. Este diseño permite obtener poder estadístico a la vez que se reducen los costes de genotipado.

4. Estudios de meta-análisis, donde se combinan resultados de asociación provenientes de diversos grupos a fin de incrementar el poder estadístico del estudio y replicar resultados a un nivel *in silico*.
5. Aplicación de estudios funcionales para determinar la causalidad de la variante genética en el desarrollo de la enfermedad.

Estos análisis arriba propuestos permiten reforzar la relación detectada entre la variante y la enfermedad (Ioannidis, 2007). La falta de confirmación de los resultados de asociación viene determinada, en gran medida, por la heterogeneidad en el riesgo entre poblaciones diferentes por la interacción de factores ambientales que, en general, no se tienen en cuenta, sumado la pérdida de información por la limitada aplicación de análisis estadísticos de combinación de polimorfismos de riesgo y protección. Así, es posible que polimorfismos inicialmente no asociados con el cáncer de mama en un estudio individual, sí que estén asociados a la enfermedad en combinación con otros cambios.

I.2.2.4. El Consorcio de Estudios de Asociación al Cáncer de Mama

Con el fin de tener un mayor poder estadístico en la detección de variantes genéticas asociadas a modestos incrementos en el riesgo a padecer cáncer de mama, se han establecido colaboraciones entre grupos de todo el mundo que han recogido, a lo largo de años de investigación, muestras de pacientes con cáncer de mama esporádico bien caracterizadas y muestras de controles de población normal. El Consorcio de Estudios de Asociación al Cáncer de Mama (BCAC, *Breast Cancer Association Consortium*) es el organismo que se ha establecido para dirigir la colaboración de 20 grupos de Europa, Norteamérica, Sureste asiático y Australia, consiguiendo un total de más de 30.000 casos de cáncer de mama y 30.000 controles de población normal. La utilización de este alto número de muestras se ha aplicado para comprobar la asociación a cáncer de mama esporádico detectada en SNPs estudiados previamente por, al menos, tres grupos independientes miembros del consorcio (Consortium, 2006).

I.3. RUTAS DE GENES DE SUSCEPTIBILIDAD A CÁNCER DE MAMA ESPORÁDICO

I.3.1. Genes implicados en la reparación del ADN: familia de genes *FANC*

La anemia de Fanconi (FA, *Fanconi anaemia*) es una enfermedad recesiva que se caracteriza por defectos congénitos como son: pigmentación anormal de la piel, fallo progresivo de la médula ósea y susceptibilidad a cáncer. Mutaciones en un grupo de genes que conforman la familia de genes *FANC* pueden causar la enfermedad a raíz de la inestabilidad cromosómica que se produce. Así, la familia de proteínas FANC ha sido identificada como un componente de la maquinaria de reparación de la rotura de la doble hebra de ADN por recombinación homóloga (D'Andrea and Grompe, 2003; Digweed *et al.*, 2002; Wang and D'Andrea, 2004). Debido a la relación existente entre la familia de proteínas FANC y el cáncer de mama, tras la identificación de *BRCA2* como *FANCD1* (Howlett *et al.*, 2002; Stewart and Elledge, 2002) y la reciente implicación de mutaciones en *FANCI* (*BRIP1*) (Bridge *et al.*, 2005; Litman *et al.*, 2005; Seal *et al.*, 2006) y en *FANCN* (*PALB2*) (Rahman *et al.*, 2007) en el riesgo a padecer cáncer de mama familiar (De Nicolo *et al.*, 2008; Tischkowitz *et al.*, 2007), la familia de genes *FANC* ha sido propuesta para el estudio de genes candidatos para cáncer de mama esporádico en la presente tesis.

Hasta ahora, han sido identificados trece genes *FANC*, que son: *FANC-A*, *-B*, *-C*, *-D1*, *-D2*, *-E*, *-F*, *-G*, *-I*, *-J*, *-L*, *-M* y *-N*. Ocho de ellos (*FANC-A*, *-B*, *-C*, *-E*, *-F*, *-G*, *-L* y *-M*) codifican proteínas que van a formar un complejo nuclear denominado complejo FANC (Gordon and Buchwald, 2003; Jacquemont and Taniguchi, 2007; Joenje *et al.*, 1997; Medhurst *et al.*, 2001). Algunas de estas proteínas actúan como elementos estructurales dentro del complejo FANC; así, por ejemplo, *FANCA* forma un dímero que, junto con *FANCG*, tienen un importante papel en el mantenimiento de la estructura del complejo (Gordon and Buchwald, 2003; Nakanishi *et al.*, 2001), mientras que *FANCF* funciona como un adaptador flexible necesario para el correcto ensamblaje del mismo (Leveille *et al.*, 2004). Otras proteínas FANC tienen funciones específicas, como *FANCL*, una ubiquitin ligasa (Meetei *et al.*, 2003; Meetei *et al.*, 2004b) cuya estabilidad y función

depende de FANCB (Fei *et al.*, 2005; Meetei *et al.*, 2004a). FANCD2, por su parte, requiere de la monoubiquitinación por FANCL para pasar de la isoforma inactiva FANCD2-S a la activa FANCD2-L (D'Andrea and Grompe, 2003; Garcia-Higuera *et al.*, 2001). FANCD2 interviene en pasos posteriores al papel del complejo FANC, al que se une a través de FANCE para la activación por monoubiquitinación arriba indicada (Gordon and Buchwald, 2003). FANCM, otra proteína perteneciente al complejo FANC, interviene en la translocación de éste a lo largo del ADN (Meetei *et al.*, 2005). Las proteínas FANCI/BRIP1 y FANCD1/PALB2, por su parte, intervienen activamente en la reparación del ADN al unirse a BRCA1 y BRCA2/FANCD1, respectivamente.

La actividad de las proteínas FANC es iniciada cuando ocurre el daño en el ADN, donde las proteínas quinasas ATM y ATR intervienen como sensores del daño (Cliby *et al.*, 1998; Morrison *et al.*, 2000). Cuando esto ocurre, dichas quinasas fosforilan a FANCD2, hecho requerido para la activación del control del ciclo celular en la fase S (Grompe, 2002; Taniguchi *et al.*, 2002b), así como para la monoubiquitinación por parte del complejo FANC (Andreassen *et al.*, 2004). ATM y ATR además fosforilan a otras proteínas del complejo FANC (Venkitaraman, 2004), necesario para la activación de FANCL y de su función ubiquitin ligasa. Después de la monoubiquitinación de FANCD2-S, FANCD2-L se transloca a la región del núcleo, ayudada por la proteína H2AX que, tras su fosforilación por parte de ATM y ATR, interviene en pasos posteriores en la correcta localización de FANCD2-L en el sitio de reparación del ADN. (Lyakhovich and Surrallés, 2007). Allí, FANCD2-L interviene en el ensamblaje de la maquinaria de reparación del ADN (Montes de Oca *et al.*, 2005; Park *et al.*, 2005), sea en respuesta al daño producido en el ADN (Hussain *et al.*, 2004) durante la fase S del ciclo normal de división celular (Rothfuss and Grompe, 2004; Taniguchi *et al.*, 2002a), o como respuesta a estrés replicativo (Howlett *et al.*, 2005) (*Figura 4*).

I.3.2. Genes implicados en el control del ciclo celular: familia de genes E2F

Los genes denominados “*gate-keepers*” intervienen en el control de la progresión del ciclo celular (Kinzler and Vogelstein, 1997) y su inactivación es suficiente para

promover el crecimiento tumoral. Así pues, estos genes parecen ser un grupo importante que podría participar en el desarrollo del cáncer de mama (Lo *et al.*, 2007).

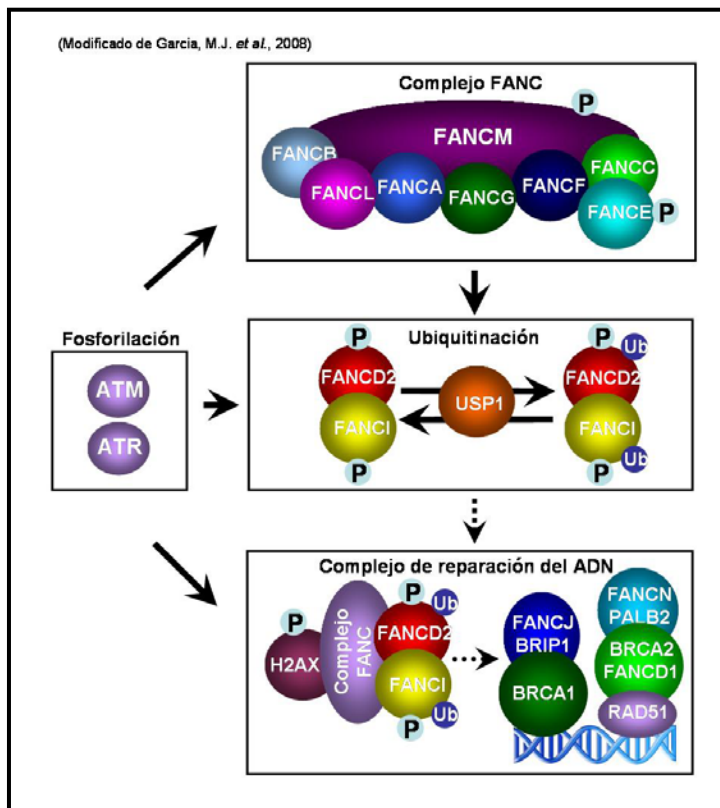


Figura 4: Funciones de la familia de proteínas FANCD. La fosforilación de las proteínas FANCM y FANCE por parte de ATM y ATR lleva a la activación de FANCL, que ubiquitina a FANCD2 y FANCI, previamente fosforiladas. Éstas se translocan al sitio de reparación del ADN y, junto con el resto de miembros del complejo FANCD, y H2AX fosforilado, permiten la activación de la función reparadora por parte de los complejos formados por BRCA1 y FANCI/BRIP1, y por BRCA2/ FANCD1, FANCN/PALB2 y RAD51.

Entre otros, se ha seleccionado a la familia de genes *E2F* que se presenta a continuación como un grupo de proteínas importante en la función de control del ciclo celular y que podría estar implicada en la aparición y desarrollo del cáncer de mama debido, entre otras cosas, a que E2F1 se ha visto relacionado con la expresión génica de *BRCA1* (Wang *et al.*, 2000a). La familia de factores de transcripción E2F está involucrada indirectamente en el control del ciclo celular. Está formada por nueve miembros: E2F-1, -2, -3a, -3b, -4, -5, -6, -7 y -8, que presentan funciones redundantes en el control del ciclo celular, pero también funciones exclusivas durante el desarrollo, la homeostasis de los tejidos y la progresión tumoral. Agrupándolos por categorías, E2F1, E2F2 y la isoforma E2F3a de *E2F3* son consideradas las proteínas E2F activadoras, indispensables en la trans-activación de genes diana que intervienen en la transición G1-S, entre otros. Éstas se unen a los promotores de los genes que activan tras la acetilación de las histonas H3 y H4 (Taubert *et al.*, 2004) y la tri-metilación de la lisina 4 de la histona H3 (H3K4me3) (Nightingale *et al.*, 2007). La isoforma E2F3b de *E2F3*, por su parte, se expresa de forma constitutiva a lo largo de todo el ciclo celular y tiene una función de represor

transcripcional, junto con E2F4, E2F5, E2F6, E2F7 y E2F8 (Attwooll *et al.*, 2004; Danielian *et al.*, 2008; Maiti *et al.*, 2005; Wu *et al.*, 2001). Las proteínas E2F1, E2F2, E2F3a, E2F3b, E2F4 y E2F5 están reguladas negativamente por las proteínas pRb, p107 (RBL1) y p130 (RBL2), denominadas comúnmente “*pocket proteins*”, que las mantienen secuestradas hasta que sus diversas funciones son requeridas (Classon and Dyson, 2001; Weintraub *et al.*, 1992). Además, forman heterodímeros con la familia de proteínas DP (*differentiation-regulated transcription factor-1 polypeptide*) compuesta por DP-1, -2 y -4. E2F6, por su parte, se une a las proteínas DP y se ha visto relacionada con proteínas del grupo “*policomb*” (PcG). E2F7 y E2F8, sin embargo, sólo forman homodímeros o heterodímeros entre ellos (van den Heuvel and Dyson, 2008) con función desconocida (Figura 5).

Si bien se ha considerado que la función fundamental de la familia E2F es la estimulación de la proliferación celular, se han descrito también otras funciones, como puede ser la inducción a apoptosis por parte de E2F1, el control de la reparación del ADN, la diferenciación de tejidos y la inducción de expresión de microRNAs por parte de E2F3 que controlan negativamente de expresión de E2F1, E2F2 y E2F3 (Polager and Ginsberg, 2008).

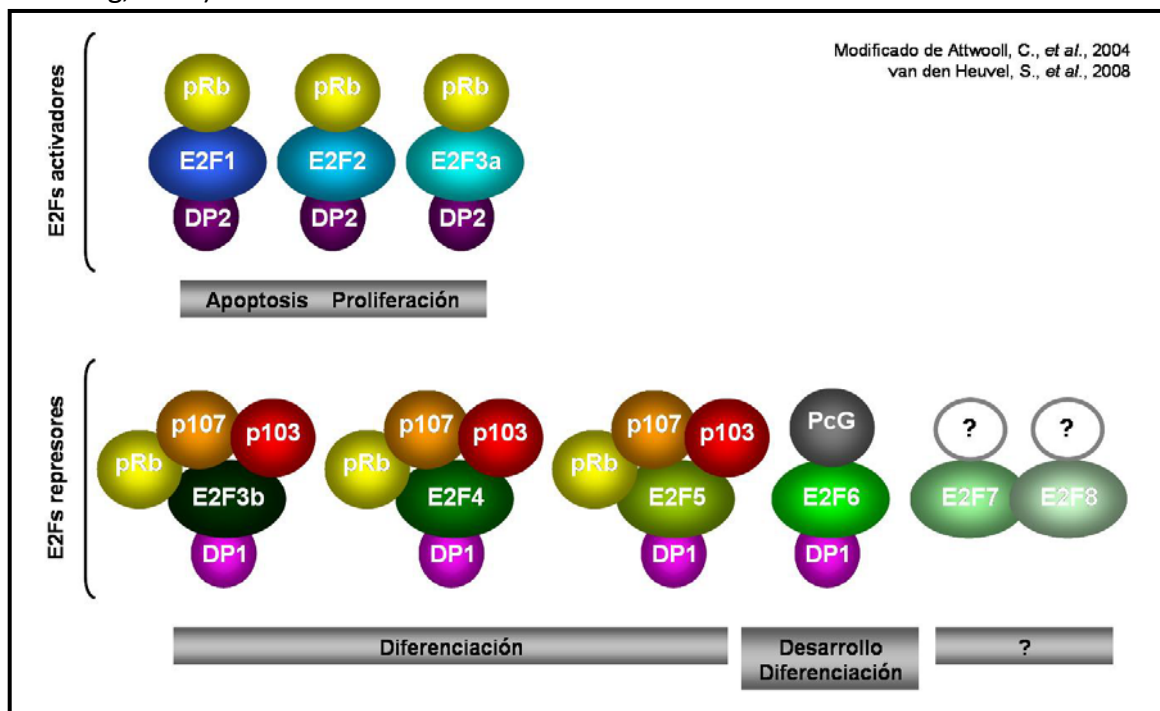


Figura 5: Representación de los subgrupos de factores de transcripción E2F activadores o represores. Se indican los complejos que forman con sus proteínas bloqueantes y las funciones que desempeñan una vez liberados.

II. OBJETIVOS

El cáncer de mama esporádico es una enfermedad común cuyos datos epidemiológicos genéticos sugieren la contribución de un número variable de genes de baja penetrancia según un modelo poligénico. Esta enfermedad compleja conformaría la hipótesis de “enfermedad común – variante común”, donde alelos de baja penetrancia en genes candidatos incrementarían la susceptibilidad a padecer la enfermedad. Así pues, los objetivos de la presente tesis doctoral fueron:

II.1. Identificación de nuevos genes de susceptibilidad a cáncer de mama esporádico:

- a. Búsqueda y selección de genes candidatos para estudios caso-control.

Seleccionamos para ello:

- a.1. El grupo de genes “*care-takers*”, o protectores del daño al ADN, reparadores del daño y reguladores de la entrada en apoptosis.
- a.2. El grupo de genes “*gate-keepers*”, o reguladores del ciclo celular y controladores, por tanto, de la proliferación celular.

- b. Búsqueda y selección de variantes genéticas:

- b.1. Variantes potencialmente funcionales localizadas en las regiones codificantes, 5’UTR, 3’UTR y promotoras.
- b.2. Variantes marcadoras o *tag*-SNPs a lo largo de todo el gen.

II.2. Validación de los SNPs que presentan asociación a cáncer de mama:

- a. Ampliación del tamaño muestral, utilización de muestras de agregación familiar del propio laboratorio o genotipación en poblaciones independientes.
- b. Análisis de la muestra estratificada según sus características fenotípicas y clínicas.
- c. Aplicación de estudios funcionales *in silico*, *in vitro* e *in vivo* sobre el efecto de los SNPs asociados.

III. MATERIALES Y MÉTODOS

III.1. MATERIAL BIOLÓGICO PARA EL ESTUDIO

III.1.1. Muestras de ADN desde sangre periférica

III.1.1.1. Obtención de muestras de sangre periférica y características generales de las muestras

El estudio de genotipado se realizó utilizando ADN genómico (ADNg) proveniente de muestras de sangre periférica de pacientes con cáncer de mama sin componente familiar y que se ha denominado cáncer de mama esporádico (CME). Además, se recogieron muestras de pacientes con cáncer de mama de agregación familiar (CMA), esto es, que al menos dos parientes de primer grado han tenido cáncer de mama y, por tanto, no pueden ser catalogadas como muestras de cáncer esporádico ni como cáncer familiar propiamente dicho. Asimismo, se utilizaron muestras de sangre periférica de personas que no padecían o habían padecido cáncer de mama y que fueron consideradas controles de población normal (CoCM). Estos tres grupos de muestras fueron remitidos a nuestro Grupo de Genética Humana desde distintos complejos hospitalarios de la Red Nacional de Salud o centros independientes, y fueron incluidos en el estudio de forma consecutiva. Las muestras de agregación familiar fueron utilizadas en algunos casos con el fin de obtener unos mejores resultados de asociación a la enfermedad debido al mayor componente genético que presentan.

Los hospitales que han colaborado en la obtención de las muestras de pacientes con cáncer de mama fueron: Hospital Monte Naranco (Oviedo), Hospital Universitario de La Paz (Madrid) y Fundación Jiménez Díaz (Madrid) llegando a un total de 547 CME. Los hospitales y centros que han colaborado para la obtención de las muestras de población normal fueron: Instituto Palacios de Salud de la Mujer (Madrid) y Fundación Jiménez Díaz (Madrid), componiendo un total de 552 CoCM (*Tabla 1*). Ambos grupos de muestras presentaban rangos de edad comparables. En aquellas ocasiones en que se requirió una comprobación del resultado de asociación obtenido, se utilizó la totalidad de las muestras que componen la colección del Grupo de Genética Humana, que asciende a 1.260 CME, 276 CMA y 1.526 CoCM y que se describen en el *Anexo 1.2*. Asimismo, se

utilizó una colección de 876 CME y 915 CoCM del grupo de H. Nevanlinna, del Departamento de Obstetricia y Ginecología de la Universidad de Helsinki, colaborador con nuestro Grupo de Genética Humana, a fin de realizar el estudio de validación en una población independiente que se indica en el apartado IV.3.3.2.

Tabla 1: Distribución de muestras CoCM y CME por hospital o centro donador.

Hospital/Centro	CoCM		CME		Total	
	N	%	N	%	N	% Total
Fundación Jiménez Díaz	92	16,7	164	30,0	256	23,3
Hospital Monte Naranco	-	-	256	46,8	256	23,3
Hospital La Paz	-	-	127	23,2	127	11,6
Instituto Palacios	460	83,3	-	-	460	41,9
Total	552	100,0	547	100,0	1.099	100,0

Los datos clínicos utilizados en el presente estudio y descritos con mayor detalle en el apartado III.1.1.5., se presentan a continuación en la *Tabla 2* desglosados según el grupo de muestras (CME o CoCM), la característica clínica o tumoral y la variable dentro de cada característica.

Tabla 2: Datos personales, clínicos y tumorales del conjunto de muestras del estudio.

Características clínicas y tumorales	CoCM (N=552)	CME (N=547)	p-valor*
	N (%)	N (%)	
Edad (años)			1,08x10⁻⁶
Mediana (rango)	53 (23-64)	58 (23-89)	
Media (D.E..)	52,0 (8,3)	57,5 (11,4)	
< Mediana CoCM	262 (47,5)	169 (30,9)	
≥ Mediana CoCM	290 (52,5)	348 (63,6)	
Desconocido	0 (0,0)	30 (5,5)	
Edad de menarquia (años)			0,200
Mediana (rango)	13 (9-18)	13 (9-19)	
Media (D.E..)	13,0 (1,5)	13,3 (1,8)	
< Mediana CoCM	157 (28,4)	140 (25,6)	
≥ Mediana CoCM	282 (51,1)	303 (55,4)	
Desconocido	113 (20,5)	104 (19,0)	
Paridad (número de hijos nacidos vivos)			5,53x10⁻⁴
Mediana (rango)	2 (0-8)	2 (0-12)	
Media (D.E..)	2,1 (1,2)	1,9 (1,5)	
Nulípara (0)	52 (9,4)	79 (14,4)	
Uníparas-Múltiparas (≥ 1)	393 (71,2)	304 (55,6)	
Desconocido	107 (19,4)	164 (30,0)	
Estado de menopausia			5,27x10⁻¹⁰
Premenopausia	221 (40,0)	116 (21,2)	
Postmenopausia	331 (60,0)	401 (73,3)	
Desconocido	0 (0,0)	30 (5,5)	
Edad de menopausia (años)			0,072
Mediana (rango)	51 (50-55)	50 (25-60)	
Media (D.E..)	52,0 (2,0)	48,7 (5,3)	
< Mediana CME	0 (0,0)	64 (11,7)	
≥ Mediana CME	5 (0,9)	81 (14,8)	
Desconocido	547 (99,1)	402 (73,5)	
Historia clínica de cáncer genérico			0,001
No	2 (0,4)	389 (71,1)	
Sí	4 (0,7)	38 (6,9)	
Desconocido	546 (98,9)	120 (21,9)	

(continuación) Característica clínica	CoCM N (%)	CME N (%)	p-valor*
Historia clínica de cáncer de mama anterior			0,036
No	4 (0,7)	406 (74,2)	
Sí	2 (0,4)	21 (3,8)	
Desconocido	546 (98,9)	120 (21,9)	
Historia familiar de cáncer de mama			-
No	-	361 (66,0)	
Sí	-	86 (15,7)	
Desconocido	-	100 (18,3)	
Estado vital			-
No	-	310 (56,7)	
Sí	-	28 (5,1)	
Desconocido	-	209 (38,2)	
Monitorización de la enfermedad (meses)			-
Media (D.E.)	-	84,4 (42,9)	
Desconocido	-	105	
Histología del tumor			-
In situ	-	50 (9,1)	
Invasivo	-	30 (67,6)	
Otros	-	25 (4,6)	
Desconocido	-	102 (18,6)	
Grado del tumor			-
Grado 1	-	102 (18,6)	
Grado > 1	-	266 (48,6)	
Desconocido	-	179 (32,7)	
Tamaño del tumor			-
≤ 2 cm	-	312 (57,0)	
> 2 cm	-	146 (26,7)	
Desconocido	-	89 (16,3)	
Metástasis de ganglios linfáticos al diagnóstico			-
Negativo	-	287 (52,5)	
Positivo	-	148 (27,1)	
Desconocido	-	112 (20,5)	
Metastasis general al momento del diagnóstico			-
No	-	441 (80,6)	
Sí	-	19 (3,5)	
Desconocido	-	87 (15,9)	
ER			-
Negativo	-	77 (14,1)	
Positivo	-	238 (43,5)	
Desconocido	-	232 (42,4)	
PR			-
Negativo	-	146 (26,7)	
Positivo	-	192 (35,1)	
Desconocido	-	209 (38,2)	

* Test exacto de Fisher, con p-valor obtenido eliminando los valores desconocidos. D.E., desviación estándar. Los resultados estadísticamente significativos están señalados en negrita rojo. Los resultados marginales se indican en negrita negro. Con un guión se indican los datos no recogidos para el grupo de muestras control.

En dicha tabla no se han presentado todos los datos clínicos tumorales obtenidos a través de los cuestionarios o por la revisión de las historias clínicas, indicados en el apartado III.1.1.5. y en el *Anexo I.3.* y el *Anexo I.4.* Esto ha sido debido a la falta de datos suficientes en algunas características clínicas y tumorales, lo que ha implicado que su inclusión en los análisis descritos en el apartado IV de Resultados no haya sido posible.

III.1.1.2. Extracción de ADN genómico desde sangre periférica

La extracción de ADNg de muestras de sangre periférica en el laboratorio de Genética Humana se ha automatizado mediante la utilización del *MagNa Pure LC DNA Isolation Kit – Large Volumen* y el robot *Magna Pure LC* (Roche, Alemania). A partir de un tubo de 10 ml de sangre periférica, se procede a la centrifugación a 3.000 rpm durante 10 minutos y a la extracción de la capa de leucocitos.

A diferencia del protocolo propuesto por Roche, se utilizan 500 µl de leucocitos, diluidos en 500 µl de suero fisiológico, en vez de 1 ml de sangre periférica, a fin de mejorar el rendimiento de extracción; a partir de este punto se ha procedido como indica el fabricante. Este tipo de extracción permite una obtención de ADNg de calidad media con una concentración promedio de 80-100 ng/µl.

III.1.1.3. Valoración de la cantidad/calidad del ADN genómico extraído

Espectrofotometría mediante radiación UV

Para estimar la cantidad de ADNg y el grado de pureza obtenidos mediante la extracción a partir de sangre periférica arriba reseñada, se valoró la absorbancia a 260 nm y a 280 nm (A_{260} y A_{280} , respectivamente) con 1 µl de ADN depositado en el lector óptico del *Nanodrop* (Thermo Fisher Scientific, EE.UU.). Como blanco del ensayo, se utilizó H₂O⁶ o TE⁶, o bien *buffer EB* (*QIAprep Miniprep Kit*, Qiagen, EE.UU.) en el caso de realizar la medición para el ADN de un plásmido.

La cantidad y, por tanto, la concentración de ADNg, se estima mediante el cálculo siguiente: $C = A_{260}/L$, donde “C” es la concentración en µg/ml, y “L” es el coeficiente de extinción molar del ADN bicatenario (ADNdc) (0,02 µg/ml).

Por otro lado, un dato importante es la pureza del ADNg obtenido, que se estima con el ratio A_{260}/A_{280} . Así pues, aquellas muestras con un ratio mayor o igual a 1,8 están

⁶ TE: Tris-EDTA, tampón de elución para ácidos nucleicos. Aumenta la estabilidad del ADN en muestras almacenadas a largo plazo a -20 °C, o en aquellas utilizadas frecuentemente, almacenadas a 4 °C o a temperatura ambiente, ya que inhibe las DNAsas presentes en la muestra. Está compuesto de Tris (tris-hidroximetil-aminometano) combinado con HCl, que lo convierte en una sal ácida, a 10 mM, que mantiene el ADN desprotonado y soluble en agua. Además tiene EDTA (ácido etilendiaminotetraacético) 1 mM, pH 8, un agente quelante de cationes divalentes, particularmente de Mg²⁺, un co-factor necesario para la función de enzimas nucleasas como las DNAsas y, por tanto, protege frente a la degradación enzimática de los ácidos nucleicos.

libres de contaminación por proteínas.

Fluorometría mediante Picogreen

La colección de muestras de casos y controles fue inicialmente cuantificada y valorada en su grado de pureza por medio del *Nanodrop*. Sin embargo, la utilización de la plataforma de genotipado *Illumina* requirió de una cuantificación más sensible de toda la colección, gracias a *PicoGreen dsDNA Quantitation Reagent Kit* (Molecular Probes, EE.UU.).

El reactivo *PicoGreen* es un agente fluorescente ultrasensible que se intercala en el ADNdc y permite una estimación de la cantidad de ADN por unidad de volumen en un rango de 25 pg/ml a 1.000 pg/ml. Las muestras son excitadas a 480 nm y la intensidad de fluorescencia de emisión se mide a 520 nm usando un espectrofluorómetro. Aporta una sensibilidad a la medición 10.000 veces mayor que aquellas obtenidas por A_{260} , cuya sensibilidad a $A_{260} = 0,1$ corresponde a 5 $\mu\text{g/ml}$ de ADN. Además, la precisión de la medición no se ve afectada por la presencia de compuestos que comúnmente contaminan las preparaciones de ácidos nucleicos, como son sales, urea, etanol y cloroformo, entre otros. Asimismo, permite la detección de ADNdc de forma exclusiva, no así la medición por absorbancia, en la que se advierte la interferencia de otros ácidos nucleicos presentes en las extracciones de ADN como son ADN de hebra sencilla (ADNcs), ARN o nucleótidos (detalle en la *Figura 6A*).

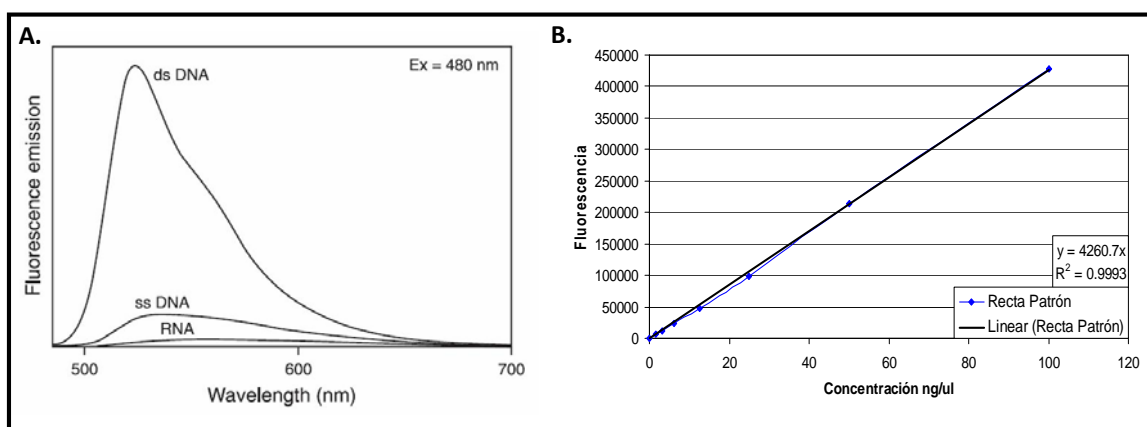


Figura 6: Lectura de PicoGreen. **A**, Ejemplo de espectro de emisión de una muestra de *PicoGreen*-ADNdc. **B**, Modelo de recta patrón utilizada para la medida de la concentración de ADNg por *PicoGreen* a partir de la inferencia de concentración.

El procedimiento requirió de la utilización de una recta patrón con ADNg a 100 ng/ μl (Clontech, EE.UU.) en cada ensayo, con la cual inferir la concentración de nuestras

muestras (detalle en la *Figura 6B*) mediante una recta de regresión lineal simple⁷. Las concentraciones utilizadas para la obtención de la recta patrón fueron las siguientes (en ng/μl): 0; 1,53; 3,13; 6,3; 12,5; 25,0; 50,0; 100,0.

El protocolo que se siguió para la medida de concentración de ADNdc fue el indicado por el fabricante. Utilizamos el robot *Beckman Coulter FX* para la dispensación de los reactivos, a fin de evitar errores, y la medida se realizó en el espectrofluorómetro *DTX 800 Multimode Detector* (Beckman Coulter, EE.UU.).

III.1.1.4. Amplificación del ADN genómico

Debido a la utilización frecuente de las muestras de ADNg del presente estudio, evitar la pérdida de éstas ha sido una cuestión clave en nuestro trabajo. Para aquellas muestras que se utilizan para genotipado se estableció un protocolo estandarizado de amplificación del ADNg utilizando el kit de amplificación *Genomiphi* (GE Healthcare, Reino Unido). Este kit utiliza la ADN polimerasa del bacteriófago Phi29, con actividad 3'→5' exonucleasa correctora de errores, para amplificar de forma exponencial el ADNdc o ADNcs con gran fiabilidad, generando fragmentos de más de 10 Kb.

Se realizó la amplificación de placas de 96 pocillos con la ayuda de la dispensación de muestras por parte del robot *Beckman Coulter FX*. La amplificación se llevó a cabo como indica el fabricante, con la salvedad de que el programa del termociclador (GeneAmp PCR system 9700, Applied Biosystems) se ha variado en el tiempo de elongación de la reacción de la ADN polimerasa como sigue:

- Incubar a 30 °C durante 2 horas. Los dNTPs serán incorporados durante este tiempo en fragmentos de gran tamaño, copias del ADN de referencia.
- Incubar a 65 °C durante 10 minutos. Se inactiva la enzima.
- Incubar a 4 °C durante 10 minutos.

El ADNg amplificado se almacena a -20 °C.

⁷ Recta de regresión lineal simple: $y = mx + b$, donde “y” es el valor de fluorescencia, “m” el coeficiente de regresión o pendiente de la recta y “b” la ordenada en el origen de la que se obtiene el valor de “x”, la variable predictora que corresponde a la concentración. El coeficiente de determinación, o cuadrado del coeficiente de correlación lineal (R^2), permite medir el correcto ajuste a la recta patrón y establecer la bondad de la predicción de la variable “x” cuando R^2 se acerca a 1.

III.1.1.5. Cuestionarios y base de datos clínicos

La recepción de cada muestra de sangre se combinó con la recepción, asimismo, de un cuestionario diseñado por el Grupo de Genética Humana del CNIO y remitido a los diferentes hospitales colaboradores, en el que el médico oncólogo ha recogido datos personales, clínicos y fenotípicos de interés para la enfermedad (ver *Anexo I.3.* para más detalle). Los datos de interés extraídos del cuestionario y añadidos a nuestro estudio fueron los siguientes: centro de procedencia, tipo de muestra (caso o control), fecha de nacimiento, sexo (los hombres se han descartado del estudio), edad de diagnóstico (del primer cáncer de mama si hubiera desarrollado otro), lugar de nacimiento (del que se deduce si es extranjera o no; aquellas pacientes extranjeras se han descartado del estudio), altura y peso 1 año antes del diagnóstico (con los que se ha calculado el IMC⁸ para estimar el sobrepeso de la paciente), edad de menarquia, número de partos, edad en el primer parto, número de abortos, edad de menopausia, menopausia natural u otras causas, consumo de alcohol, consumo de tabaco, sobrepeso, toma de anticonceptivos, tratamiento de estimulación ovárica, tratamiento hormonal sustitutivo, historia personal de cáncer, historia familiar de cáncer, tipo histológico del tumor, grado de diferenciación del tumor, TNM (del que se deduce un tamaño de tumor menor o igual, o mayor a 2 cm, según T1; afectación del ganglios linfáticos, según N1; metástasis al diagnóstico, según M1), y estado de los marcadores inmunohistoquímicos RE (receptor de estrógenos), RP (receptor de progesterona), Ki-67 (antígeno identificado por el anticuerpo monoclonal Ki-67), p53 y HER2/neu (ErbB2).

Estos datos fueron completados bajo revisión de las historias clínicas, para las muestras provenientes de los centros hospitalarios La Paz (Madrid) y Monte Naranco (Oviedo). Ello contribuyó a completar datos clínicos y tumorales, como son: historia personal y familiar de cáncer de mama de primer y segundo grado de parentesco, tipo de tumor, grado de diferenciación del tumor, TNM y los marcadores inmunohistoquímicos ya descritos. Además, se obtuvieron nuevos datos de seguimiento del estado de la paciente, como son: estado vital, número de meses de seguimiento, presencia de cáncer

⁸ IMC: Índice de masa corporal, calculado según: $IMC = \text{peso (Kg)} / (\text{altura (m)})^2$. Los rangos normales corresponden a un ratio de 18-25, por debajo se considera delgadez y por encima sobrepeso. La obesidad se alcanza sobrepasando un valor de 30, mientras que la delgadez severa se estima por debajo de 16.

bilateral, presencia de metástasis y marcadores inmunohistoquímicos de RA (receptor de andrógenos) y Bcl-2, entre otros (ver *Anexo I.4.*).

Asimismo, para la obtención de datos de controles de población normal, se utilizó un cuestionario simplificado en el que se recogieron datos relevantes para el estudio del cáncer de mama y comunes a los encuestados para las pacientes con cáncer de mama: edad al realizar el estudio, edad de menarquia, número de partos, edad en el primer parto, número de abortos, edad de menopausia, menopausia natural u otras causas, consumo de alcohol, consumo de tabaco, sobrepeso, toma de anticonceptivos, tratamiento de estimulación ovárica, tratamiento hormonal sustitutivo, historia personal de cáncer e historia familiar de cáncer (ver *Anexo I.3.* para más detalle).

Con todo ello se hizo una base de datos clínica que comprende los 1.260 CME, 276 CMA y 1.526 CoCM mencionados anteriormente, y que incluye los datos que se especifican en el *Anexo I.4.* Los detalles de los datos fenotípicos y clínicos del conjunto de muestras utilizadas en el presente trabajo se presentaron en la *Tabla 2.*

Todos los participantes en este estudio son de origen español y caucásico. Además, todos han dado su consentimiento informado, aprobado por el Comité de Ética del Hospital La Paz.

III.1.2. Muestras de tejido embebido en parafina

III.1.2.1. Obtención de muestras de tejido embebido en parafina

En determinados estudios funcionales de inmunohistoquímica y PCR cuantitativa a tiempo real fueron necesarios bloques de tejido tumoral y/o normal embebidos en parafina. Estos bloques fueron requeridos a los centros hospitalarios de La Paz (Madrid) y Monte Naranco (Oviedo), recuperando un total de 4 muestras con componente tumoral y 37 muestras de tejido mamario normal, según se especifica en la *Tabla 3.*

Tabla 3: Recuento de los tejidos parafinados obtenidos por hospital.

Hospital	N tumor	N Normal
Hospital Monte Naranco	2	12
Hospital La Paz	2	25
Total	4	37

III.1.2.2. Extracción de ARN total desde tejidos embebidos en parafina

La extracción de ARN total desde tejidos embebidos en parafina, utilizado en ensayos de PCR cuantitativa a tiempo real, comprende los pasos que se describen a continuación y descritos por Tarón y colaboradores (Taron *et al.*, 2004).

a. Pre-tratamiento de los portaobjetos:

Se utilizan cristales en el primer paso del proceso de extracción de ARN. Cada portaobjetos lleva tantos cortes de una muestra como caben en su superficie. Así pues, es necesaria la limpieza total de ésta de RNAsas que puedan degradar el ARN obtenido en los últimos pasos de la extracción. Para ello se siguen los siguientes pasos:

- Lavar los portaobjetos con *RNAse ZAP* (Ambion, EE.UU.) durante 5 minutos.
- Lavar por dos veces con H_2O , durante 5 minutos cada una, y dejar secar al aire.

b. Cortes del tejido parafinado:

Para evitar contaminación con RNAsas, es necesario limpiar el microtomo con etanol al 70% y cambiar las cuchillas cada nuevo corte. Además, el baño histológico debe tener H_2O a 42 °C sin gelatina. Por último, las secciones de tejido deben tener un grosor de 5 µm. No se deben guardar los cortes durante más de un mes.

c. Desparafinación del tejido parafinado:

Se libera el tejido de la parafina en la que está embebido. Para ello, se realiza lo siguiente:

- Emplazar los cristales a 65 °C durante 1 hora, a fin de derretir la parafina y que se retire del cristal. El tejido queda pegado al portaobjetos.
- Tratar con xilol precalentado a 65 °C durante 10 minutos para retirar los restos de parafina y deshidratar el tejido. Dejar secar al aire.

d. Digestión del tejido:

- Recuperar el tejido del cristal al añadirle 1 gota (~10 µl) de tampón de digestión⁹ y completar hasta 200 µl para la digestión a 60 °C en agitación durante toda la noche.
- Segunda digestión con 10 µl de proteinasa K (20 mg/ml), incubar a 60 °C en agitación durante 15-20 minutos.

⁹ Tampón de digestión: Se utiliza con el fin de liberar los ácidos nucleicos que contienen las células. Está compuesto por 10 mM Tris-HCl pH 8.0, 0,1 mM EDTA pH 8.0 (actuarán de estabilizadores de los ácidos nucleicos), SDS 2% (dodecilsulfato sódico, un compuesto tensoactivo iónico que degrada las membranas y rompe los enlaces covalentes de las proteínas, desnaturalizándolas) y 500 mg/ml de proteinasa K (una proteasa endolítica que degrada las proteínas).

e. Precipitación y aislamiento de los ácidos nucleicos:

Para ello se siguen los siguientes pasos:

- Separación de los ácidos nucleicos de las proteínas al utilizar 300 µl de PCI (25 fenol:24 cloroformo:1 ácido isoamílico, 10 mM Tris, pH 8, 1 mM EDTA, de Sigma, EE.UU.) por muestra, emplazar en hielo durante 10 minutos y centrifugar a máxima velocidad durante 10-15 minutos a 4 °C.

- Recoger la fase acuosa en la que se encuentran los ácidos nucleicos, añadir 20 µl de acetato sódico, 2.5 µl de glucógeno y 300 µl de isopropanol (fríos) y centrifugar a máxima velocidad durante 10 minutos a 4 °C. Ello va a contribuir a la precipitación de los ácidos nucleicos en la mezcla.

- Eliminar el sobrenadante, añadir 150 µl de etanol al 70% y centrifugar a máxima velocidad durante 5 minutos. Permitirá la eliminación de las sales acumuladas en la mezcla a lo largo del proceso de precipitación de los ácidos nucleicos.

- Eliminar el sobrenadante y dejar secar en estufa.

- Resuspender en 53 µl de H₂Odd DEPC¹⁰ libre de RNAsas. Incubar a 65 °C durante 5 minutos para la total hidratación de los ácidos nucleicos.

f. Eliminación del ADN y purificación del ARN.

Se realiza un tratamiento con el kit *DNA-free DNase Treatment & Removal Reagents* (Ambion, EE.UU.), con las variaciones que se especifican a continuación, a fin de obtener el ARN purificado.

- Añadir 1 µl de DNasa y 6 µl de *DNase buffer 10x* e incubar a 37 °C durante 20 minutos.

- Añadir 1 µl de DNasa e incubar a 37 °C durante 20 minutos más.

- Inactivar la DNasa al añadir 12 µl de *DNase inactivation reagent 10x*, incubar durante 2 minutos a temperatura ambiente y centrifugar a 10.000 rpm durante 1 minuto. Este tampón es un agente bifásico que va a permitir la separación de una fase acuosa en la que se encontrará el ARN, que será la que se recoja y guarde a -20 °C, y una fase orgánica que atraparà a la DNasa.

¹⁰ DEPC: El dietil pirocarbonato es un agente alquilante y un fuerte inhibidor de RNAsas. Se utiliza a una concentración de 0,1% en H₂Odd.

III.1.3. Muestras de líneas celulares

III.1.3.1. Líneas celulares utilizadas.

Las líneas celulares utilizadas en este trabajo fueron HeLa y MCF-7. La línea celular HeLa, proveniente de un carcinoma cervical, es comúnmente utilizada en la investigación biomédica. Ello es debido a su inmortalidad, que reporta un fácil mantenimiento en cultivo con un ilimitado número de divisiones mientras existan condiciones generales de supervivencia celular (Masters, 2002). Por otra parte, la línea celular MCF-7 proviene de un adenocarcinoma mamario. Se trata de una línea celular bien caracterizada, que presenta RE positivos, de modo que es una línea útil para modelar el cáncer de mama *in vitro* y su relación con el estrógeno. Además, son células positivas para RP, citoqueratina, E-cadherina y receptores del factor de crecimiento epidérmico, mientras que son negativas para receptores del factor de crecimiento de fibroblástico, vimentina y muchas de la metaloproteasas de la matriz celular (Soule *et al.*, 1973).

III.1.3.2. Extracción de ADN genómico desde líneas celulares

La extracción de ADNg de líneas celulares se realizó como sigue:

- Lavar las células del medio de cultivo mediante: centrifugación de las células en un tubo a 1.500 rpm durante 5 minutos, eliminación del medio sobrenadante, adición de 1 ml de PBS¹¹ 1x, resuspensión de las células, nueva centrifugación a 1.500 rpm durante 5 minutos, posterior eliminación del sobrenadante, y obtención del sedimento de células limpias.
- Añadir 1 ml de *DNAzol Genomic DNA Isolation Reagent* (Molecular Research Center, EE.UU.) y resuspender. Este agente va a permitir la rotura total de las células y la liberación de los ácidos nucleicos.
- Añadir 1 ml más de *DNAzol* e incubar en balancín a temperatura ambiente al menos durante 24 horas.

¹¹ PBS: Tampón fosfato salino (phosphate buffered saline) es una solución comúnmente utilizada en biología molecular, con múltiples aplicaciones pues es isotónica y no tóxica para las células. El PBS está compuesto de NaCl 137 mM, fosfatos (Na₂HPO₄ y KH₂PO₄) 10 mM, KCl 2,7 mM, y pH 7,4.

- Una vez están las células totalmente digeridas, centrifugar a 4.000 rpm durante 3 minutos.

- Recoger todo el sobrenadante, donde se encuentran los ácidos nucleicos, añadir 300 µl de etanol absoluto por cada 1 ml de *DNAzol* que se puso, a fin de lavar los ácidos nucleicos, y agitar por balanceo hasta ver la hebra de ADN.

- Recoger la hebra de ADN con una pipeta Pasteur con un extremo de gancho y lavar con etanol al 70% para re-hidratar el ADN.

- Secar al aire, hasta la completa evaporación del etanol.

- Re-hidratar la hebra de ADN en H₂O_d o TE (en cantidad variable según la estimación de obtención de ADN, se varía entre 30-50 µl) y agitar a 37 °C durante toda la noche. Guardar a 4 °C, o bien a -20 °C para tiempos prolongados.

III.1.3.3. Extracción de ARN total desde líneas celulares

La extracción de ARN total de líneas celulares se realizó como se describe a continuación:

a. Rotura de las células:

- Sobre el sedimento de células limpias de medio por el lavado con PBS 1x descrito arriba, añadir 1 ml de *Trizol* (Invitrogen, Reino Unido) por cada 3,5 cm de diámetro de la placa de cultivo, no por número de células, y recoger las células resuspendiendo bien. Este agente permite la rotura total de las células y la liberación de proteínas y ácidos nucleicos que pueden ser aislados.

- 1 ml de *Trizol* por cada pocillo de una placa de 6 pocillos (3,5 cm de diámetro, 10 cm² de superficie).

- 400 µl de *Trizol* por cada pocillo de una placa de 12 pocillos (4 cm² de superficie).

- Pasar el volumen de células resuspendidas a un tubo de 1,5 ml y congelar a -80 °C al menos durante toda la noche. Pueden guardarse de esta manera durante meses.

b. Separación de fases:

- Descongelar las células en *Trizol* a temperatura ambiente y esperar 10 minutos adicionales después de estar descongeladas para que se rompan bien.

· Añadir 200 µl de cloroformo por cada 1 ml de *Trizol*, a fin de que la fase orgánica atrape a las proteínas. Agitar durante 15 segundos por inversión, incubar a temperatura ambiente durante 2-3 minutos y centrifugar a 12.000 g durante 15 minutos a 4 °C.

· El ARN queda en la fase acuosa y se transfiere a un tubo nuevo.

c. Precipitación del ARN:

· Añadir 500 µl de isopropanol frío por cada 1 ml de *Trizol* utilizado. Agitar mecánicamente, incubar a temperatura ambiente durante 10 minutos y centrifugar a 12.000 g durante 10 minutos a 4 °C. El ARN queda en el fondo del tubo.

d. Lavado y resuspensión del ARN:

· Retirar el sobrenadante y lavar con 1 ml de etanol al 75% (guardado a -20 °C), resuspender por agitación y pipeta suavemente. En este paso se puede guardar 4 °C durante 1 semana o a -20 °C durante 1 mes. Centrifugar a 7.500 g durante 5 minutos a 4 °C. Repetir el lavado si es necesario.

· Secar al aire, pero sin dejar que se seque totalmente.

· Resuspender con H₂O_d DEPC, e incubar a 55-60 °C durante 10 minutos para terminar de resuspender el ARN.

e. Testado del ARN:

Se estudió el estado del ARN extraído mediante un gel de electroforesis de agarosa *low EEO* (electroósmosis) (Pronadisa, España) en TBE¹². Verificación de la integridad del ARN extraído al ver las dos bandas discretas de ARN ribosómico 28S y 18S, mientras que no se observa un ARN degradado en forma de “smear”. Además, se realiza una lectura de su concentración en el *Nanodrop*.

¹² TBE: Tris-Borato-EDTA. Es un tampón amortiguador comúnmente utilizado en todas las aplicaciones de electroforesis de ADN, tanto para geles de agarosa como de acrilamida. En general este tampón ofrece mejor resolución cuando se están testando fragmentos de 0.1-3 Kb, y es mejor para electroforesis con voltajes >150V por su baja conductividad.

III.2. IDENTIFICACIÓN Y SELECCIÓN DE GENES Y POLIMORFISMOS

III.2.1. Selección de genes de interés para el cáncer de mama

La selección de los genes que hemos estudiado en este trabajo se ha basado en la revisión bibliográfica a lo largo del tiempo de realización del proyecto. Así, se ha prestado atención a aquellos genes que codifican proteínas relacionadas de alguna manera con el cáncer de mama, tanto porque tienen una función implicada en procesos carcinogénicos en general, como por su relación con el desarrollo de tejidos epiteliales como es el tejido mamario.

III.2.2. Bases de datos y selección de SNPs

Se ha utilizado una serie de aplicaciones bioinformáticas y bases de datos de carácter público que han reportado información a fin de seleccionar los SNPs en los genes elegidos para el presente estudio.

III.2.2.1. Criterios de selección de los SNPs

Los criterios de selección de los SNPs para el presente estudio fueron los siguientes:

- SNPs que presentan una MAF descrita en una población caucásica europea o de origen europeo mayor o igual al 10%. Se ha permitido una excepción a este requisito, con una MAF entre 5% y 10%, a aquellos SNPs que se encuentran en regiones codificantes y que suelen tener una menor frecuencia que los demás, a fin de no desechar variantes putativas funcionales en el gen y que pudieran estar implicadas en la enfermedad.

- Se ha dado prioridad a aquellos SNPs que pueden tener una función en la expresión del gen y que se encuentran no sólo en las regiones codificantes, sino también en zonas que pueden intervenir en la regulación de la expresión génica, como son las uniones exón-intrón (que actúan en el “*splicing*” del ARNm, ARN mensajero), la región

promotora (tomada como aquella que llega hasta 1 Kb corriente arriba con respecto al codón de iniciación, donde los SNPs pueden estar interrumpiendo la unión de factores de transcripción a sus secuencias consenso, o creando nuevas dianas para éstos), las regiones activadoras o represoras del “*splicing*” presentes en intrones y exones, y las regiones no traducidas a proteína (UTR, *untranslated region*) en los extremos 5’ y 3’ (5’UTR y 3’UTR, respectivamente) del gen.

- Se han considerado además los *tag*-SNPs o SNPs descritos como marcadores moleculares genéticos que definen los distintos bloques de DL, a fin de recoger toda la información de variabilidad genética de una región del genoma o de un gen.

- Los SNPs no se deben encontrar en un pseudogén, una secuencia repetitiva, una secuencia similar a otras del genoma, o demasiado cerca de otros SNPs pues haría que el diseño de genotipación fallara.

- Los SNPs pre-seleccionados por los criterios anteriores deben estar conservados a lo largo de la evolución de mamíferos, dando idea de la importancia que dichas variantes tienen en la función del gen.

III.2.2.2. Recogida de datos de SNPs de una región genómica

Se realizó una búsqueda de toda la información de variabilidad genética en la región génica que engloba a los genes de interés, incluyendo los extremos 5’ y 3’ hasta 1 Kb de distancia, aproximadamente. Para ello se usaron las bases de datos siguientes: NCBI (<http://www.ncbi.nlm.nih.gov/>), Single Nucleotide Polymorphism Database (<http://www.ncbi.nlm.nih.gov/SNP/>), Ensembl (<http://www.ensembl.org/index.html>) y HapMap (<http://www.hapmap.org/index.html.en>).

Ello permitió la descarga de los datos siguientes: número de identificación del SNP (codificación rs, *reference SNP*), posición cromosómica y posición dentro del gen (en pares de bases, pb), localización génica (exón, intrón, extremos 5’UTR y 3’UTR, y extremos 5’ y 3’ fuera de la región génica), tipo y posición en el gen del cambio peptídico sinónimo o no sinónimo (en pb), tipo de cambio nucleotídico, definición del alelo menor y MAF en las poblaciones caucásicas europeas PDR90, AFD-EUR Panel North America o

III. MATERIALES Y MÉTODOS

III.2. Identificación y selección de genes y polimorfismos

CSHL-HapMap-CEU¹³. Un ejemplo de los datos extraídos usando estas bases de datos de SNPs es el que se presenta en la *Figura 7*. La realización de un *BLAST* o comparación de la secuencia de los SNPs de interés con todo el genoma permitió determinar su singularidad (<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>).

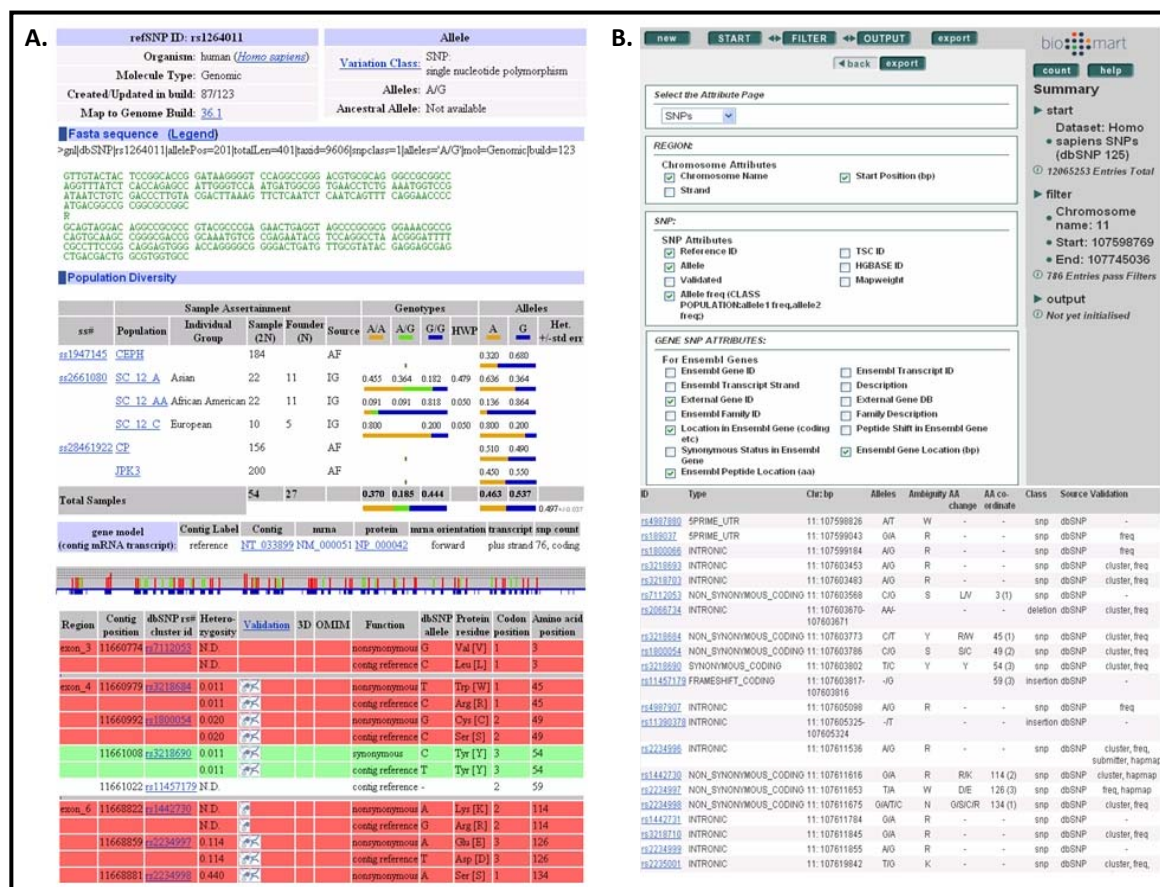


Figura 7: Ejemplo de búsqueda y descarga de datos de variabilidad genética. Se han utilizado: **A**, las bases de datos de NCBI; **B**, las bases de datos de Ensembl.

III.2.2.3. Definición de bloques de desequilibrio de ligamiento y tag-SNPs

Se utilizó *Haploview* v4.1 (<http://www.broad.mit.edu/mpg/haploview/>) (Barrett *et al.*, 2005), una aplicación de *HapMap*, para la obtención de información de los bloques

¹³ **PDR90:** Es la definición de la población caucásica estadounidense estudiada por el *National Institute of Health (NIH)* de EE.UU., encuadrada dentro del proyecto *Polymorphism Discovery Resource (PDR)*. Está compuesta por 90 individuos, según: 24 personas de origen europeo, 24 personas de origen africano, 12 personas de origen mejicano, 6 nativos norteamericanos y 24 personas de origen asiático.

AFD-EUR Panel North America: Es la definición de la población caucásica compuesta por 23 individuos estadounidenses no relacionados y de ascendencia europea.

CSHL-HapMap: Este panel está compuesto por 270 personas: 30 tríos de población Yoruba de Ibadan, Nigeria (YRI); 45 individuos no relacionados de población asiática de Pekín; 45 individuos no relacionados de población asiática de Tokio (CHB+JPT); 30 tríos *CEPH* (*Centre d'Etude du Polymorphisme Humain*), una colección de residentes de Utah con ascendencia del norte y este de Europa (CEU).

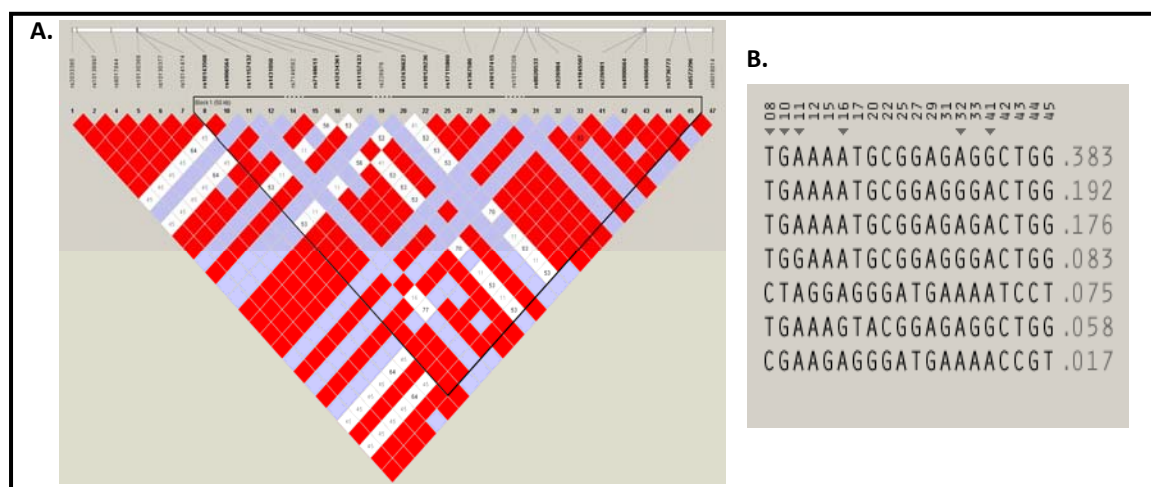
de DL de la región genómica de interés. Los bloques de DL definen la relación entre dos alelos que presenta una mayor frecuencia que la que ocurriría por mero azar, debido a que estos alelos están cercanos físicamente en el cromosoma y es infrecuente un evento de recombinación que los separe (ver *Figura 8A*). Para esta estimación, el programa *Haploview* tiene en cuenta la distancia cromosómica entre los dos loci estudiados y estima las dos medidas más frecuentes de DL: D' y r^2 , ambas relacionadas con D , la unidad básica de DL. D es una medida de desviación de la frecuencia de haplotipos del estado de equilibrio, definida según: $D = f(AB) - (f(A) f(B))$, donde “f” es la frecuencia alélica, “A” es el alelo mayor del SNP 1, y “B” es el alelo mayor del SNP 2. Un *haplotipo* se define como un grupo de marcadores genéticos que están ligados, se presentan en un sólo cromosoma y tienden a heredarse conjuntamente; algunos haplotipos pueden estar en DL. Cuando D es mayor que 0, existe DL. D' , por otra parte, es el ratio absoluto de D comparado con su máximo valor, D_{max} , cuando $D \geq 0$, o bien con su mínimo valor, D_{min} , cuando $D < 0$, una medida estandarizada de DL que establece rangos entre 0 y 1. $D' = 0$ implica que los dos loci están completamente desligados, $D' = 1$ supone que los loci son idénticos, mientras que aquellos valores cercanos a 1 sugieren que los dos loci están cercanos en el cromosoma y, por tanto, pertenecen al mismo bloque de DL. El valor r^2 es el coeficiente de determinación estadístico, una medida de correlación entre un par de variables que se obtiene según la ecuación siguiente: $r^2 = D / (f(A) f(a) f(B) f(b))$. Esta medida es de particular importancia en el mapeo genético, puesto que está inversamente relacionada con el tamaño muestral requerido en un estudio de asociación. Así pues, conocido el genotipo y los valores de frecuencia alélica de un SNP determinado, para predecir de forma directa el genotipo de un SNP desconocido es necesario un valor de $r^2 = 1$ y un mayor tamaño muestral. D' y r^2 se relacionan entre ellos y con las frecuencias alélicas según la ecuación: $r^2 = (D')^2 \times (f(a) f(B) / (f(A) f(b)))$, donde D' aporta información de la extensión física del bloque de DL al estimar el límite máximo de r^2 (Wang *et al.*, 2005).

Otra aplicación de *Haploview* es la herramienta *tagger-pairwise*, que permite definir los *tag*-SNPs. Así, con un número reducido de SNPs se puede definir la variabilidad genética de una región y evitar, por tanto, el gasto innecesario de recursos en el

III. MATERIALES Y MÉTODOS

III.2. Identificación y selección de genes y polimorfismos

genotipado de los SNPs a los que, *per-sé*, los *tag*-SNPs describen. Un ejemplo de estimación de *tag*-SNPs es el que se presenta en la *Figura 8B*.



III.2.2.4. Búsqueda de SNPs con funcionalidad en el gen

La herramienta bioinformática *PupaSNP*, denominado posteriormente como *PupaSuite* v1.0 (<http://pupasuite.bioinfo.cipf.es/>) (Conde *et al.*, 2006; Conde *et al.*, 2004) se utilizó para la determinación de la función de los SNPs en el gen. En esta aplicación se incluyen los datos obtenidos de: *Ensembl* v44, *dbSNP* 126, *HapMap* release #21, *Transfac Professional* v8.3, *Match* v2.2, *Haploview* v3.32, *SNPEffect* v3.0, *Jaspar*, *GeneID* y *miRanda*. Se analizaron SNPs codificantes, aquellos SNPs que se encontraban en la región de unión intrón-exón y que pudieran afectar a la función de “*splicing*” que esta unión realiza (Hull *et al.*, 2007), SNPs exónicos e intrónicos que pueden interferir en la regulación del “*splicing*”, al localizarse en zonas potenciadoras o inhibidoras del mismo (ElSharawy *et al.*, 2006), y aquellos que regulan la función de microRNAs (Saunders *et al.*, 2007). Un ejemplo de la ventana de datos de *PupaSuite* se presenta en la *Figura 9A*.

Otra importante herramienta que se ha utilizado con el fin de seleccionar aquellos SNPs más interesantes fue *Functional Element SNP Database* (<http://combio.kribb.re.kr/FESD/>), con la que se obtuvieron datos acerca de la localización de los SNPs en el gen y su posible funcionalidad (*Figura 9B*).

La herramienta *Transfac Professional* v8.3, dentro del paquete *rVista* v2.0 (<http://rvista.dcode.org>) (Loots and Ovcharenko, 2004), fue aplicada con el fin de detectar secuencias consenso de unión de factores de transcripción en la región promotora de los genes a estudio. Un ejemplo de la visualización de esta herramienta se presenta en la *Figura 9C*.

Las herramientas *Pfam* (<http://www.sanger.ac.uk/Software/Pfam/>) y *EMBL-EBI* (<http://www.ebi.ac.uk/>), fueron utilizadas para definir las zonas de la proteína con una evidente actividad, de modo que SNPs que se localizaran en estas zonas serían putativos de ejercer un efecto importante en la funcionalidad de la proteína codificada (*Figura 9D*).

El análisis de conservación filogenética se realizó utilizando *ECR Browser* (<http://ecrbrowser.dcode.org>) (Ovcharenko *et al.*, 2004). Esta aplicación permitió establecer la filogenia a lo largo de la evolución de los mamíferos: opossum (*M. domestica*), rata (*R. norvegicus*), ratón (*M. musculus*), perro (*C. familiares*), chimpancé (*P. troglodytes*) y humano (*H. sapiens*). Aunque otras especies no mamíferas también pudieron ser incluidas, restringimos el estudio a las seis mencionadas. La representación de esta herramienta se presenta en la *Figura 22A*. Además, *Tcoffee* (<http://igs-server.cnrs-mrs.fr/Tcoffee/tcoffee.cgi/index.cgi>) (Notredame *et al.*, 2000) fue utilizada para construir árboles filogenéticos de los organismos seleccionados para los genes indicados, además de aportar información a nivel de secuencia de los cambios acontecidos a lo largo de la evolución en aquellas regiones que incluyen SNPs de interés, como se muestra en la *Figura 23* y en la *Figura 24*.

III. MATERIALES Y MÉTODOS

III.2. Identificación y selección de genes y polimorfismos

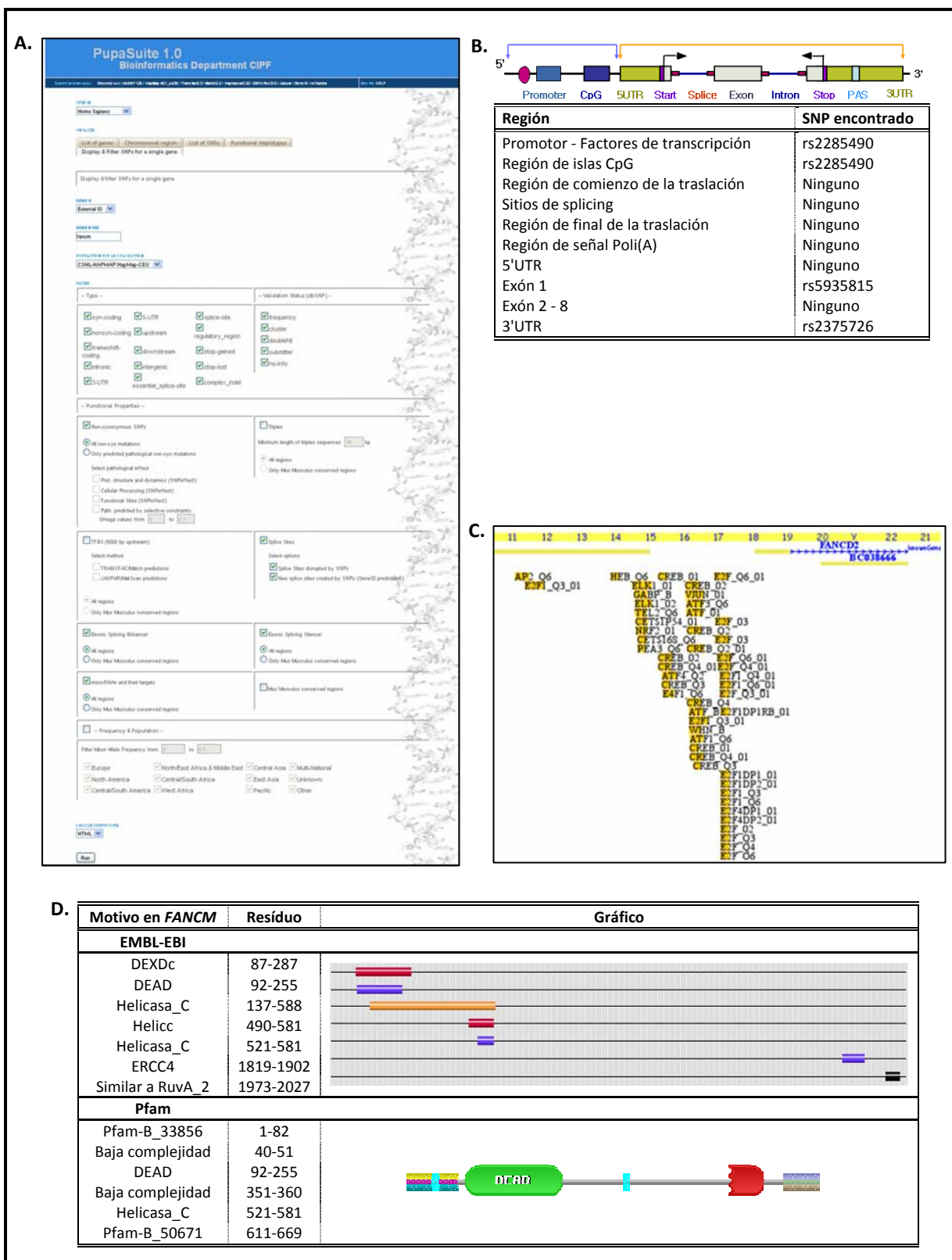


Figura 9: Herramientas de búsqueda de SNPs funcionales en el gen. A, Visualización de los datos recogidos por la herramienta *PupaSuite*. **B,** Esquema de los elementos analizados por *Functional Element SNP Database* y tabla resumen de los SNPs en cada elemento. **C,** Descripción *in silico* de los sitios consenso de unión de factores de transcripción en la región promotora de un gen según *Transfac*. **D,** Caracterización de los motivos funcionales de la proteína según *EMBL-EBI* y *Pfam*.

III.3. GENOTIPADO DE LOS SNPs SELECCIONADOS

III.3.1. Control de calidad de los ensayos de genotipado

La calidad de los ensayos descritos a continuación se verificó incorporando un control negativo (muestra sin ADN) y al menos 2 duplicados por cada placa de 96 muestras de ADN. Además, se incluyeron duplicados entre ensayos para comprobar la reproducibilidad de la técnica. Se obtuvo un 100% de concordancia en todos los SNPs estudiados.

III.3.2. PCR

La reacción en cadena de la polimerasa o PCR (*polimerase chain reaction*) es la tecnología sobre la que se basan todas las técnicas de genotipado que se van a exponer a continuación y que, por su generalizada utilización en Biología Molecular, no se va a detallar. Sin embargo, es de interés comentar el diseño de los oligonucleótidos que se realizó en todo el proyecto, según los pasos siguientes:

- Selección de la secuencia que se quiere amplificar y realización de un *BLAST* a fin de confirmar que los oligonucleótidos no van a amplificar otras regiones del genoma.
- Utilización del programa de diseño de oligonucleótidos *Primer 3* (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) (Rozen and Skaletsky, 2000) que sigue los parámetros de diseño que se especifican a continuación: tamaño del oligo entre 18 y 27 pb, temperatura de “*melting*” (T_m)¹⁴ entre 57 °C y 63 °C, y porcentaje de GC entre 20% y 80%, entre otros.

Los oligonucleótidos utilizados en los diferentes ensayos de PCR y sus condiciones específicas de amplificación se presentan en el *Anexo I.5*.

¹⁴ T_m : Es la temperatura de *melting* (fusión), o disociación del ADNdc, a la cual el 50% de las bases se encuentran apareadas, mientras que el otro 50% se encuentra en forma de cadena sencilla. Esta temperatura depende de la longitud de la cadena, del porcentaje de G-C y la concentración de sales en la solución. Se puede calcular de forma manual, según la siguiente fórmula, para temperaturas mayores de 55°C: $T_m = 66 + 0,55 (\% \text{ G+C}) - 675 / L$, donde “% G+C” es el contenido de GC (%), y “L” es la longitud del oligonucleótido.

La purificación del producto de PCR se hace necesaria para no interferir en la reacción de secuenciación. Para ello es necesario eliminar el excedente de dNTPs y oligonucleótidos no incorporados en la secuencia, así como las enzimas y sales que componen la mezcla de reacción. Se han seguido 3 procedimientos distintos en función de la aplicación posterior del producto de PCR:

1. Purificación por JETQUICK PCR Purification Spin Kit (Genomed), según las instrucciones del fabricante. Este procedimiento utiliza columnas donde el ADN se une, se limpia de los componentes indicados anteriormente y, finalmente, se eluye.

2. Purificación por ExoSAP-IT™ (USB, EE.UU.), compuesto por la Exonucleasa I (*ExoI*), y por la fosfatasa alcalina de gamba ártica (*SAP, shrimp alkaline phosphatase*), que degradan el excedente de dNTPs y oligonucleótidos de la reacción de PCR. Para ello se utilizan tan sólo 3 µl de producto que se van a introducir en una mezcla de SAP buffer 10x, 10 U/µl de *ExoI*, 1 U/µl de *SAP* y H₂O hasta 6 µl de volumen final de reacción. Se dispone en un programa de 37 °C durante 15 minutos y 72 °C durante 15 minutos. A diferencia de la purificación antes mencionada, se utiliza muy poco producto, lo que facilita la posterior utilización del producto de PCR para otros fines.

3. Purificación por deshidratación, considerado el método clásico. Para realizarlo, se siguen los pasos que se especifican a continuación:

- Añadir 250 µl de etanol absoluto frío (guardado a -20 °C) a la PCR o producto de ligación de vector más inserto. Mezclar y dejar precipitar el ADN a -20 °C durante 15 minutos.

- Centrifugar durante 20 minutos a máxima velocidad, decantar el sobrenadante.

- Añadir 250 µl de etanol al 70% frío (guardado a -20 °C). Mezclar y dejar precipitar el ADN a -20 °C durante 10 minutos.

- Centrifugar durante 20 minutos a máxima velocidad y volver a decantar el sobrenadante.

- Dejar evaporar durante 1 hora o más, hasta que el etanol se haya evaporado totalmente.

- Diluir el producto de ADN purificado en el volumen de H₂O o TE deseado, con un mínimo de 5 µl.

III.3.3. TaqMan

III.3.3.1. Principios metodológicos de la discriminación alélica por *TaqMan*

El genotipado de SNPs mediante la técnica *TaqMan*[®] *SNP Genotyping Assays* (Applied Biosystems, EE.UU.) (Livak *et al.*, 1995), está basado en la detección directa de un producto de PCR como consecuencia de la liberación de un fluorocromo. Para ello, la actividad exonucleasa 5'→3' de la enzima ADN polimerasa y el principio físico conocido como "Transferencia de energía por resonancia de fluorescencia" (FRET, *fluorescence resonance energy transference*), son claves en este proceso.

El ensayo *TaqMan* hace uso de dos sondas específicamente diseñadas para reconocer los dos alelos a discriminar. Cada sonda consiste en un oligonucleótido que porta:

- Una molécula fluorogénica en el extremo 5' de cada sonda: fluorocromo 6-FAM (6-carboxifluoresceína) para la detección del alelo 1 y fluorocromo VIC para la detección del alelo 2.
- Un grupo receptor de fluorescencia (NFQ, *non-fluorescence quencher*) en el extremo 3' de la sonda, cuya longitud de onda de excitación coincide con la longitud de onda de emisión del fluorocromo del extremo 5'. De esta forma, la proximidad del grupo receptor al fluorocromo del extremo 5' conduce a una supresión de la fluorescencia de este último, siguiendo el principio de transferencia de energía por resonancia de fluorescencia.
- Una molécula de unión al surco menor del ADN (MGBP, *minor groove binding protein*) en el extremo 3', que permite una mayor eficacia en la discriminación alélica al incrementar las diferencias de T_m entre los oligonucleótidos unidos y no unidos al ADN.

El proceso de PCR en los ensayos de discriminación alélica es similar a una reacción de PCR común, pero se diferencia de ésta en que se unen tanto los cebadores encargados de amplificar los fragmentos como las sondas que diferencian ambos alelos, que se unen a una de las hebras de ADN desnaturalizado. Durante la extensión, la ADN polimerasa se encuentra en esa hebra de ADN con la sonda o sondas hibridadas (dependiendo del genotipo homocigoto o heterocigoto), lo que se traduce en la

activación de la actividad exonucleasa 5'→3', la digestión de la sonda o sondas y la separación del fluorocromo de su inhibidor (grupo receptor no fluorescente), como se señala en la *Figura 10A*. Así pues, a lo largo de los ciclos se van generando mayor número de moléculas de fluorocromo libre. Las muestras heterocigotas generarán señal de ambos fluorocromos (VIC y FAM) mientras que las muestras homocigotas producirán señal sólo de uno de los dos fluorocromos (VIC o FAM) (*Figura 10B*).

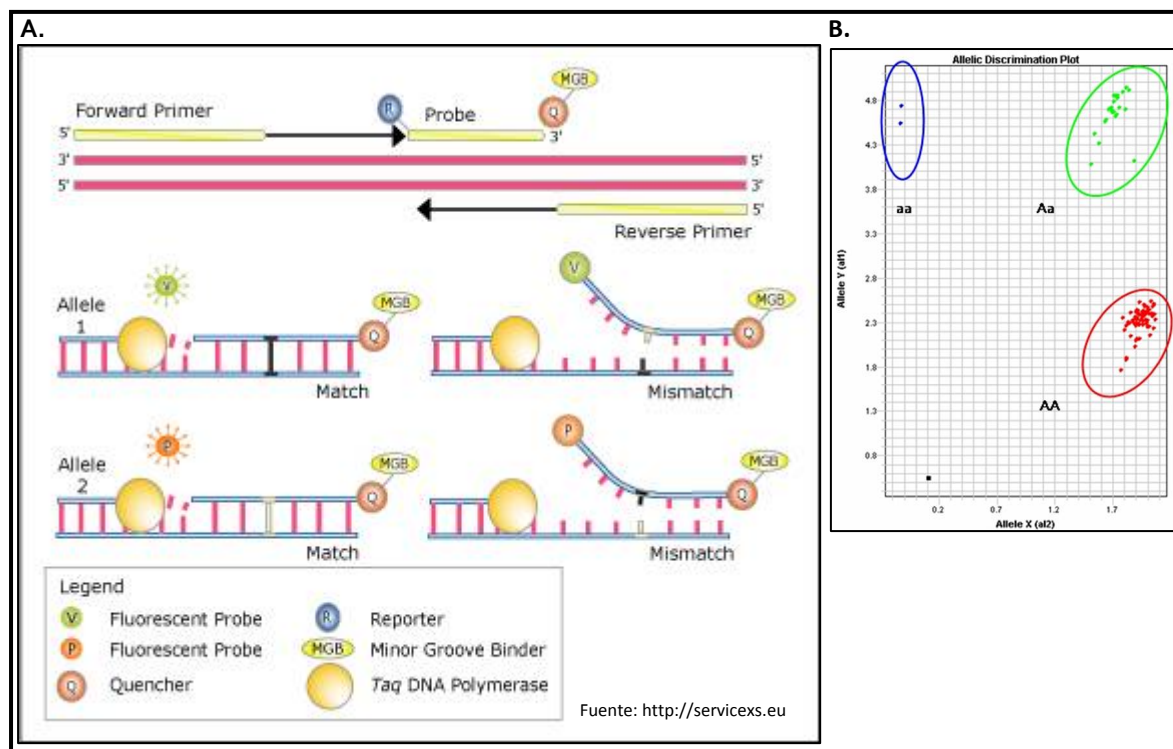


Figura 10: Principio de acción de la discriminación alélica por TaqMan. **A**, Esquema del principio de acción de la tecnología TaqMan de discriminación alélica dividido en 4 pasos que incluyen: anillamiento de la sonda a la secuencia diana, extensión e hidrólisis de la sonda, liberación de la sonda del efecto del grupo receptor de fluorescencia y transferencia de energía por resonancia de fluorescencia. **B**, Visualización de los tres grupos de genotipos obtenidos tras un genotipado TaqMan: homocigoto mayoritario (AA, alelo 2, VIC, en rojo), heterocigoto (Aa, alelos 1 y 2, FAM y VIC, en verde) y homocigoto minoritario (aa, alelo 1, FAM, en azul).

III.3.3.2. Ensayo de discriminación alélica por TaqMan

La mezcla de reacción utilizada para el genotipado de SNPs (con un volumen final de 5 µl) consta de los siguientes reactivos:

- Una mezcla 40x que contiene los oligonucleótidos no marcados utilizados para la PCR (36 µM), así como las sondas TaqMan (marcadas con FAM y VIC) para la detección de los 2 alelos (8 µM cada una). Todos estos componentes están diseñados específicamente para cada uno de los ensayos de discriminación alélica. El diseño Assay-

by-design requiere de los pasos siguientes: selección del SNP y la secuencia diana, informe a Applied Biosystems de la sonda que se requiere, medida de la calidad de la secuencia (no hay secuencias repetitivas, pseudogenes o SNPs muy cercanos al SNP interrogado) y elaboración de la sonda. En todos los casos se requirió una sonda a una escala *V-Scale* para 1.000 reacciones, utilizando 5 µl de mezcla en placas de 384 pocillos. Las sondas *Applied-on-Demand* están pre-diseñadas y testadas por la compañía.

- Tampón de reacción *TaqMan Universal PCR Master Mix 2x* (UMM), con todos los elementos necesarios para la realización de la PCR: AmpliTaq Gold DNA polimerasa, dNTPs, tampón específico y MgCl₂. Incluye además la enzima AmpErasa para evitar la formación de amplificadores inespecíficos a temperatura ambiente: inactiva a AmpliTaq hasta que recibe un choque de calor, se desnaturaliza y comienza la amplificación específica.

- 4-10 ng de ADN (2,25 µl de ADN a 5 ng/µl), que fue previamente dispensado mediante el robot *Beckman Coulter FX* en placas de 384 pocillos y evaporado a temperatura ambiente durante 24 horas.

- H₂O hasta completar el volumen final de la reacción.

La mezcla de reacción se ha ajustado a 4 µl totales en vez de los 5 µl propuestos en aquellas ocasiones en las que los fluorocromos presentan una buena señal, optimizando la utilización de la sonda con la que se pueden genotipar hasta 1.500 muestras. El dispensado de la mezcla fue realizado en la mayoría de las ocasiones con la ayuda del robot *Beckman Coulter FX*, con el fin de agilizar el genotipado, requiriendo un 8% de exceso en previsión de un gasto mayor que con un dispensado manual.

La amplificación se llevó a cabo en termocicladores (GeneAmp PCR system 9700, Applied Biosystems), según: desnaturalización de la AmpErasa a 58 °C durante 2 minutos; desnaturalización previa a 95 °C durante 10 minutos; 40 ciclos de desnaturalización a 95 °C durante 15 segundos, y anillamiento y elongación a 60 °C durante 1 minuto.

Cada ensayo fue analizado mediante el sistema *ABI PRISM Sequence Detection System 7900 HT* (SDS v2.1, Applied Biosystems, EE.UU.). Según los datos de fluorescencia obtenidos para las 2 sondas (VIC y FAM), se procedió a la definición de los 3 grupos (homocigotos para el alelo 1, homocigotos para el alelo 2 y heterocigoto) (según *Figura*

10B), mediante el software *SDS* v2.1, que permite generar estos grupos con un 95 % de intervalo de confianza.

En el *Anexo I.6.* se presentan las sondas *TaqMan* que se han diseñado para este trabajo.

La eficiencia de la sonda para discriminar los dos posibles alelos de un locus fue comprobada en un grupo de 24 muestras de población control y 1 control negativo (blanco sin muestra de ADN). Cuando estos ensayos no permitían diferenciar adecuadamente los grupos genotípicos en las condiciones de temperatura inicialmente establecidas, se realizó una segunda prueba con condiciones menos restrictivas, reduciendo la temperatura de anillamiento y extensión a 58 °C. El fallo de reacción de PCR en estas condiciones implicaba descartar definitivamente el ensayo como método de discriminación alélica. Siguiendo esta estrategia, se rechazó aproximadamente un 5-10% de las sondas que se diseñaron.

III.3.4. Amplifluor

III.3.4.1. Principios metodológicos de la discriminación alélica por *Amplifluor*

El genotipado de SNPs mediante la técnica *Amplifluor*® *SNPs HT Genotyping System* (Chemicon, EE.UU.) se realizó en aquellos casos en los que *TaqMan* o *iPLEX* fallaron debido a problemas en el ensayo. Está basado en el mismo principio que *TaqMan*, en la detección directa de un producto de PCR tras la liberación de un fluorocromo. El principio FRET se aplica de forma similar en este proceso de discriminación alélica, dado que la molécula fluorogénica está de nuevo acoplada a un complejo aceptor “*quencher*” DABSYL (4-(dimetilamino)azo benceno - ácido sulfónico) que va a limitar su actividad a un momento concreto de la reacción debido a la transferencia de energía que se produce entre ambos. Este ensayo incluye:

- Dos oligonucleótidos marcados universales (UniPrimers) que llevan una secuencia en el extremo 3' que es complementaria al extremo 5' de los oligonucleótidos específicos de alelo, y una secuencia en el extremo 5' que va a formar una estructura secundaria en forma de horquilla. Los fluorocromos FAM y JOE (6-carboxi-4',5'-dicloro-2',7'-dimetoxifluoresceína) se encuentran localizados en la primera base del extremo 5'

del UniPrimer, mientras que el grupo receptor de fluorescencia está unido al nucleótido que es complementario a esta base en 5'. De esta forma, de igual modo que pasaba con *TaqMan*, la proximidad del grupo receptor al fluorocromo del extremo 5' impide la fluorescencia de este último.

- Tres oligonucleótidos no marcados, uno reverso y dos directos específicos del SNP y que son diseñados para la amplificación de la región del SNP. Éstos últimos llevan acoplados en su extremo 5' la secuencia señal reconocida por los UniPrimers.

El proceso de PCR y discriminación alélica es similar al que ocurre en *TaqMan*. Durante la hibridación de los oligonucleótidos específicos de alelo, el acoplamiento de los oligonucleótidos marcados y la posterior amplificación, la estructura de horquilla se destruye y la transferencia de energía entre el fluorocromo y el “quencher” se interrumpe, liberando un mayor número de moléculas de fluorocromo libre en cada ciclo (Figura 11). De forma similar a como ocurría con *TaqMan*, las muestras heterocigotas generarán señal de ambos fluorocromos (FAM y JOE), mientras que las muestras homocigotas producirán señal sólo de uno de los dos fluorocromos (FAM o JOE).

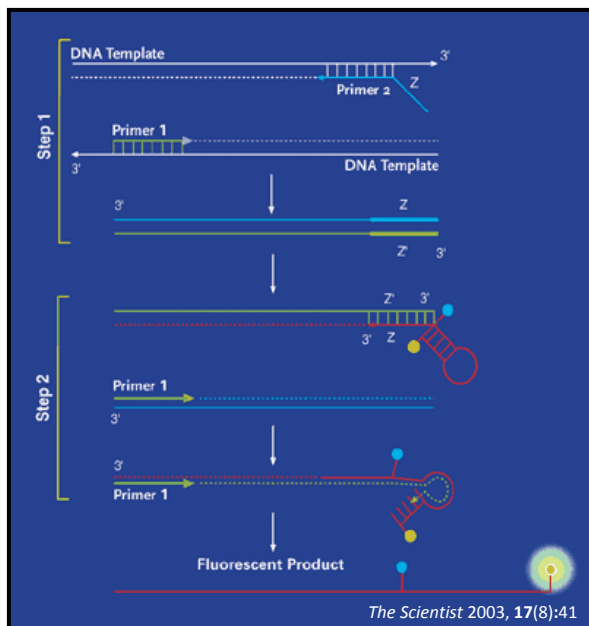


Figura 11: Esquema del principio de acción de la tecnología Amplifluor de discriminación alélica. Presenta dos fases: Step 1: amplificación de la secuencia de ADN con los oligonucleótidos específicos de alelo; Step 2: marcaje y fluorescencia de cada alelo.

III.3.4.2. Ensayo de discriminación alélica por *Amplifluor*

Debido a que los fluorocromos están acoplados a oligonucleótidos universales y la discriminación alélica es llevada a cabo por oligonucleótidos normales, pudiendo utilizar

los UniPrimers en multitud de ensayos, se trata de una metodología más económica que *TaqMan*.

La mezcla de reacción utilizada para el genotipado de SNPs (con un volumen final de 5 µl) consta de los siguientes reactivos:

- Una mezcla 10x que contiene los tres oligonucleótidos no marcados utilizados para la PCR. El diseño de estos oligos se llevó a cabo según *Amplifluor AssayArchitect Software*.

- Una mezcla 20x de los UniPrimers FAM y JOE y de los oligonucleótidos de amplificación de la región que incluye el SNP de interés.

- Los elementos necesarios para la PCR, como son: Taq Titanium DNA polimerasa (que contiene un anticuerpo que impide la acción de la enzima a temperatura ambiente, mientras que permite su actuación una vez comienza la PCR y se desnaturaliza) 5 U/µl, dNTPs 2,5 mM y tampón S-Plus 10x.

- 4-10 ng de ADNg (2,25 µl de ADN a 5 ng/µl), que fue previamente dispensado mediante el robot *Beckman Coulter FX* en placas de 384 pocillos y evaporado a temperatura ambiente durante 24 horas.

- H₂O hasta completar el volumen final de la reacción.

La amplificación se llevó a cabo en termocicladores (*GeneAmp PCR system 9700*, Applied Biosystems), siguiendo el siguiente programa:

- Desnaturalización del anticuerpo a 96 °C durante 4 minutos.

- Primer paso de la amplificación con repetición de 20 ciclos: desnaturalización a 96 °C durante 10-15 segundos, anillamiento a 55-60 °C durante 5 segundos y elongación a 72 °C durante 10 segundos.

- Segundo paso de la amplificación con repetición de 22-24 ciclos: desnaturalización a 96 °C durante 10-15 segundos, anillamiento a 55 °C durante 20 segundos y elongación a 72 °C durante 40 segundos.

- Elongación final a 72 °C durante 3 minutos.

Cada ensayo fue analizado mediante el sistema *7900 HT Sequence Detection System* (SDS v2.1, Applied Biosystems, EE.UU.) y, como ocurría con *TaqMan*, se definieron los 3 grupos de genotipos (homocigotos para el alelo 1, homocigotos para el

alelo 2 y heterocigotos). En el *Anexo I.7.* se indican los oligonucleótidos utilizados en el ensayo de discriminación alélica con *Amplifluor*.

III.3.5. dHPLC

III.3.5.1. *dHPLC* como alternativa a *TaqMan* y *Amplifluor*

El uso de la cromatografía líquida desnaturalizante de alto rendimiento (*dHPLC*, *denaturing high pressured liquid chromatography*) como herramienta de discriminación alélica fue considerado únicamente en aquellas situaciones en las que los ensayos mediante sondas *TaqMan* o *Amplifluor* no permitieron discriminar correctamente los distintos grupos genotípicos o cuando el diseño de las sondas no era posible. Esto último pudo deberse a que la secuencia cercana al polimorfismo de interés impedía la generación de las sondas o a que la presencia de polimorfismos próximos al polimorfismo de interés hacía imposible diseñar sondas que permitiesen diferenciar apropiadamente ambos alelos. Además, la aplicación más importante de esta tecnología es la detección de variantes no descritas o la identificación de todos los cambios presentes en el fragmento de interés.

III.3.5.2. Principios metodológicos del *dHPLC*

Esta técnica permite detectar sustituciones de un único cambio de nucleótido al retener de forma distinta en una matriz no porosa los homodúplex y heterodúplex (fragmentos de ADN que muestran o no una complementariedad absoluta, respectivamente) de un fragmento amplificado mediante PCR.

El sistema utilizado, *WAVE Nucleic Acid Fragment Analysis System* (Transgenomic, Reino Unido), consiste en una columna formada por cuentas de poliestireno-divinilbenceno. Este copolímero, que está alquilado con dos cadenas carbonadas de 18 átomos (C-18), se caracteriza por ser electrostáticamente neutro e hidrofóbico, por lo que no retiene los ácidos nucleicos. Para la absorción del ADN, se utiliza acetato de trietilamonio (TEAA), reactivo hidrofílico y positivamente cargado que interactúa con la carga negativa de los grupos fosfato del ADN, mientras que los grupos hidrofóbicos del

TEAA interactúan con las cadenas C-18 de las cuentas de poliestireno, sirviendo como conexión entre esta matriz y el ADN.

La columna descrita previamente es lo que se conoce como fase estacionaria. La fase móvil consiste en una combinación de tampones que separan el ADN de la columna. La solución utilizada para fluir los fragmentos amplificados es un tampón 0,1 mM TEAA y 25% acetonitrilo, de forma que a medida que la concentración de acetonitrilo en la columna va aumentando, la interacción entre el TEAA y los fragmentos amplificados se reduce hasta liberar el ADN. Este proceso es lo que se conoce como cromatografía líquida de intercambio iónico en fase reversa (más detalles en la *Figura 12*).

III.3.5.3. Estudio de genotipación a través del *dHPLC*

Para detectar variantes se realizaron PCRs para las que los oligos diseñados se presentan en el *Anexo I.5.*, y los fragmentos amplificados fueron dispensados en una placa de 96 pocillos y sometidos a un ciclo de temperaturas que favorece la formación de heterodúplex (95 °C durante 15 minutos e incubación a 65 °C durante 1 hora) previamente al paso por la columna de *dHPLC* (ver *Figura 12*). De esta forma, las muestras de individuos heterocigotos para un polimorfismo presentan un ratio 1:1 de ADN normal:ADN mutado. Las condiciones de elución para cada fragmento se presentan, asimismo, en el *Anexo I.5*.

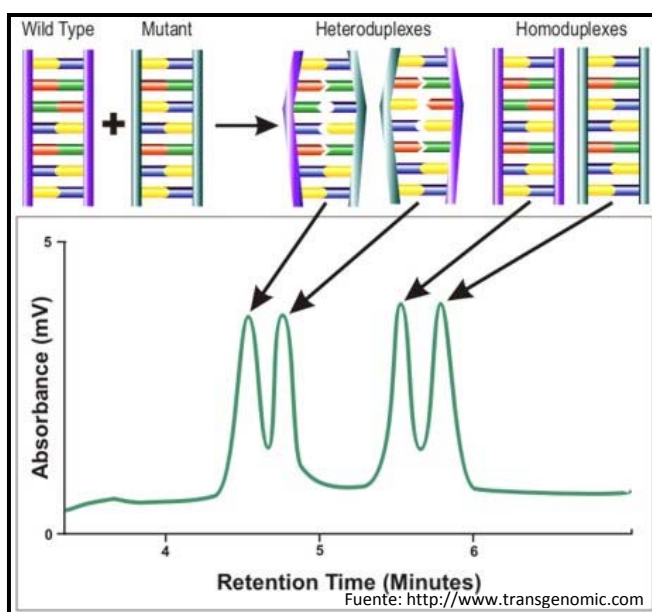


Figura 12: Principio metodológico del *dHPLC*. Formación de heterodúplex y perfil cromatográfico.

El software del sistema *WAVEMAKER Navigator Software* v1.0 (Transgenomic, Reino Unido) analiza la secuencia y predice las condiciones desnaturalizantes (temperaturas de 52-75 °C) apropiadas para el análisis mutacional del fragmento, si bien la temperatura de análisis puede ser modificada para detectar con mayor eficacia la presencia de la variante seleccionada, para lo que se observa más específicamente las condiciones de desnaturalización de la secuencia próxima a la variante.

Todas las variantes detectadas por cromatografía líquida se definieron por técnicas de secuenciación. Una vez se asoció cada patrón cromatográfico a la presencia de la variante en homocigosis o en heterocigosis, se analizaron las muestras con la intención de asignar cada muestra a cada uno de los genotipos.

III.3.6. Secuenciación

Los productos de PCR purificados fueron secuenciados de forma automática utilizando terminadores fluorescentes Big Dye™ (Applied Biosystems, EE.UU.) y un secuenciador multicapilar *ABI 3700*.

Inicialmente, se realiza una reacción de secuenciación cíclica de 20 µl de volumen final, que contiene 10 pM del oligonucleótido elegido, 5% de DMSO, 4 µl de Big Dye terminador y 4-12 ng del producto purificado en función del tamaño del fragmento a secuenciar. La reacción se lleva a cabo en un termociclador *Gene Amp PCR System 9700* (Perkin Elmer, EE.UU.) y consta de 10 ciclos en los que la muestra es sometida a desnaturalización a 95 °C durante 10 segundos, anillamiento de los oligonucleótidos a 50 °C durante 5 segundos y elongación a 60 °C durante 4 minutos.

El producto final de esta reacción es purificado por ultrafiltración en columnas de Sephadex G-50 (Millipore), para ser analizado posteriormente mediante el secuenciador multicapilar *ABI 3700* (Applied Biosystems, EE.UU.) y visualizado gracias al software *Chromas* v1.6.

III.3.7. MassArray iPLEX Gold

III.3.7.1. Principios metodológicos de la discriminación alélica por iPLEX

La tecnología de *MassARRAY® iPLEX Gold* (Sequenom, San Diego, EE.UU.) deriva de la conocida como ensayo *MassEXTEND homogéneo* o *hME* (*homogeneous MassEXTEND*), en el que el oligonucleótido *MassEXTEND* se une a un sitio adyacente al polimorfismo de interés y, mediante una reacción de PCR, se produce la elongación del oligonucleótido sobre el SNP. El producto resultante de la extensión del oligonucleótido, entre 1 y 4 bases más largo que el primer original, se analiza por *MALDI-TOF* (desorción-ionización mediante láser asistida por matriz-tiempo de vuelo, *Matrix-Assisted Laser Desorption/Ionization - Time-Of-Flight*) y es empleado para determinar la secuencia de nucleótidos en el sitio polimórfico en un proceso de minisequenciación. Con *hME* el número máximo de SNPs por ensayo es de 15. Sin embargo, en el caso de *iPLEX Gold*, todas las reacciones terminan tras una extensión de una base (*SBE, single base extension*) y se incorporan terminadores con masa modificada (ver *Tabla 4*), lo que permite genotipar hasta 36 SNPs por ensayo, aunque este número depende de los SNPs candidatos para hacer el diseño, la presencia de motivos en la secuencia y las interacciones entre oligonucleótidos.

III.3.7.2. Ensayo de discriminación alélica por iPLEX

En el ensayo *iPLEX* se realizan los pasos que se describen a continuación (ver *Figura 13* para más detalle):

a. Amplificación de la región de interés:

Se realiza la amplificación de todas las regiones de interés incluyendo los SNPs seleccionados mediante una primera reacción de PCR multiplex utilizando oligonucleótidos normales de amplificación. La mezcla de reacción utilizada para la PCR de amplificación (con un volumen final de 5 µl) consta de los siguientes reactivos:

- Una mezcla de oligonucleótidos a 500 nM cada uno.
- Los elementos necesarios para la PCR: HotStart Taq ADN polimerasa (que contiene un anticuerpo que impide la acción de la enzima a temperatura ambiente,

mientras que permite su actuación una vez comienza la PCR y se desnaturaliza) 5 U/μl, dNTPs 25 mM, MgCl₂ 25 mM y tampón de PCR 10x.

- 5-10 ng de ADN (2,25 μl de ADN a 5 ng/μl), que fue previamente dispensado mediante el robot *Beckman Coulter FX* en placas de 384 pocillos y evaporado a temperatura ambiente durante 24 horas.

- H₂O hasta completar el volumen final de la reacción.

La amplificación se llevó a cabo en termocicladores (*GeneAmp PCR system 9700*, Applied Biosystems), siguiendo el siguiente programa:

- Desnaturalización del anticuerpo a 94 °C durante 15 minutos.
- Amplificación con repetición de 45 ciclos: desnaturalización a 94 °C durante 20 segundos, anillamiento a 56 °C durante 30 segundos y elongación a 72 °C durante 1 minuto.
- Elongación final a 72 °C durante 3 minutos.

b. Defosforilación:

La defosforilación de cualquier nucleótido residual de la amplificación previene su incorporación en el paso de extensión de oligonucleótidos, mediante SAP. Para ello se añadieron 2 μl de la mezcla de SAP (SAP buffer 10x, 1 U/μl de SAP y H₂O) a 5 μl del producto de PCR, e incubando a 37 °C durante 20 minutos y 85 °C durante 5 minutos.

c. Incorporación y extensión de oligonucleótidos:

Se procede a la incorporación a la reacción de los oligonucleótidos anexos a los SNPs y a la extensión de una sola base con ddNTPs¹⁵ con modificación de masa mediante una PCR especial. El resultado es un conjunto de fragmentos (dos por cada SNP correspondientes a cada alelo) que serán discriminados por su diferente masa. La mezcla de reacción utilizada para la reacción *iPLEX* (con un volumen final de 2 μl) consta de los siguientes reactivos:

- Una mezcla de oligonucleótidos de extensión ajustando sus concentraciones de 7-14 μM cada uno dependiendo de su masa: aquellos con menor masa deberán estar a una concentración final en el volumen de reacción de 0,625 μM, mientras que los de mayor masa estarán a 1,25 μM.
- Los elementos necesarios para la PCR: *iPLEX* buffer 10x y la enzima de *iPLEX*.

¹⁵ ddNTPs: Dideoxirribonucleótidos trifosfatos, que han perdido el elemento –OH en los carbonos 3'-C y 2'-C de la desoxirribosa.

- El producto del paso anterior ya purificado.
- H₂O hasta completar el volumen final de la reacción.

El ciclo de reacción de *iPLEX* se llevó a cabo en termocicladores (*GeneAmp PCR system 9700*, Applied Biosystems), siguiendo el siguiente programa:

- Desnaturalización a 94 °C durante 30 segundos.
- Amplificación con repetición de 5 ciclos: anillamiento a 52 °C durante 5 segundos y elongación a 80 °C durante 5 segundos.
- Amplificación con repetición de 40 ciclos: desnaturalización a 94 °C durante 5 segundos, anillamiento a 52 °C durante 5 segundos y elongación a 80 °C durante 5 segundos.
- Elongación final a 72 °C durante 3 minutos.

d. Eliminación de sales:

La eliminación de las sales del producto de reacción mediante el uso de una resina es un paso necesario para evitar el ruido en el análisis espectrométrico que las sales propias de la reacción de PCR podrían introducir. Para ello, los productos de la reacción *iPLEX* se diluyen en una mezcla de 16 µl de H₂O y 6 mg de resina y se centrifugan a 4.000 rpm.

e. Dispensación en el *SpectroCHIP bioarray*:

Se utilizó un robot nanodispensador para la dispensación de los productos de *iPLEX* de cada muestra en el *SpectroCHIP bioarray*.

f. Análisis MALDI-TOF:

Finalmente, el *SpectroCHIP bioarray* se dispone dentro del escáner *MassARRAY MALDI-TOF*, donde un láser incide sobre cada producto correspondiente a cada muestra para todos los SNPs incluidos en el estudio, de modo que vuelen en un tubo sin aire. Los fragmentos más ligeros alcanzan el receptor antes que los más pesados, permitiendo su identificación gracias a la utilización del *MassARRAY Workstation software v3.3*.

En el *Anexo 1.8*. se indican los oligonucleótidos utilizados en el ensayo de discriminación alélica con *iPLEX*.

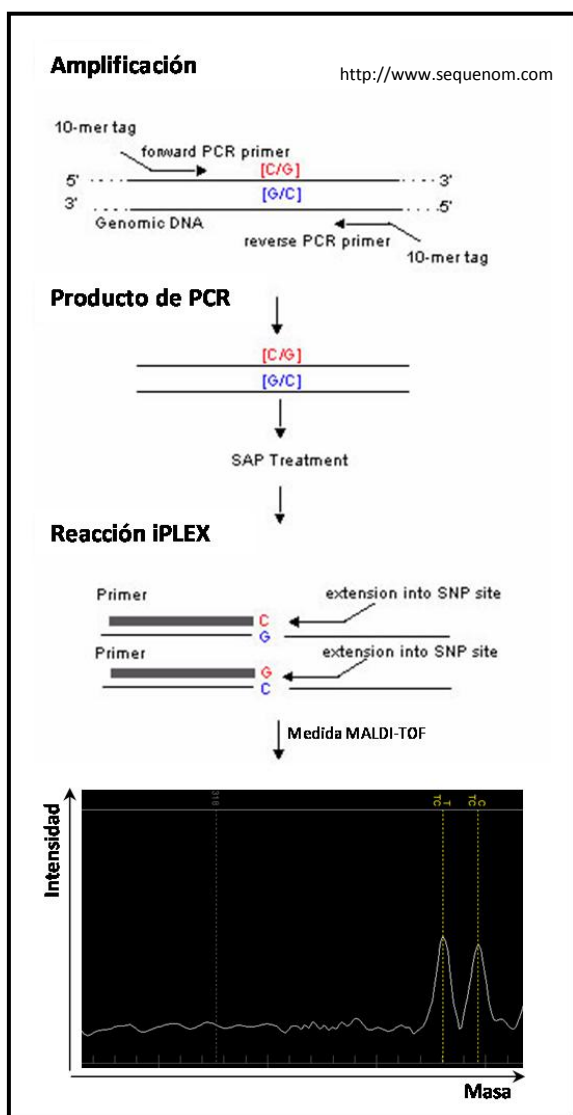


Tabla 4: Diferencias de masa entre los productos de iPLEX.

Terminador	A	C	G	T
A	0	-24	16	55,9
C	24	0	40	79,9
G	-16	-40	0	39,9
T	-55,9	-79,9	-39,9	0

Figura 13: Principio metodológico del iPLEX. Amplificación de la región de interés, defosforilación del producto de PCR, reacción iPLEX y lectura por MALDI-TOF.

III.3.8. Illumina GoldenGate

III.3.8.1. Principios metodológicos de la discriminación alélica por *Illumina GoldenGate*

La plataforma *Illumina GoldenGate*® (Illumina, San Diego, EE.UU.) (Fan *et al.*, 2003; Shen *et al.*, 2005) constituye un sistema de genotipado masivo que permite estudiar de 384 a 1.536 polimorfismos de forma simultánea en 96 muestras, generando aproximadamente 150.000 genotipos por ensayo. Cada punto de genotipado está formado por una fibra conductora de luz que está químicamente modificada para crear un pocillo microscópico al final de cada una de ellas. En este pocillo se inserta una cuenta

que contiene a su vez múltiples copias de un único oligonucleótido o “secuencia señal”, útil para estudiar una única variante.

III.3.8.2. Ensayo de discriminación alélica por *Illumina GoldenGate*

El ensayo se inicia con la incorporación del ADN_g en forma de ADN_cs a un soporte sólido. Sobre este ADN se produce el anillamiento de 3 cebadores diferentes:

- 2 oligonucleótidos específicos de los alelos de interés, que constan de un extremo 3' complementario a la secuencia próxima al polimorfismo y que se diferencian entre sí exclusivamente en la última base (correspondiente al SNP), y un extremo 5' que incorpora una secuencia para la unión de cebadores universales (P1, ACTTCGTCAGTAACGGAC, y P2, GAGTCGAGGTCATATCGT, cada una de ellas asociada a un alelo distinto). Cada oligonucleótido está marcado con un fluorocromo (cianinas Cy3 y Cy5).

- 1 oligonucleótido complementario a la zona P3 del cebador específico de locus que, desde el extremo 5' al 3', consta de: una secuencia específica de locus, una región que actúa como “secuencia señal” al ser complementaria a la existente en cada uno de los diferentes puntos de genotipado, y una secuencia para el anillamiento de cebadores universales (GTCTGCCTATAGTGAGTC).

Una vez anillados estos oligonucleótidos a su correspondiente región, una ADN polimerasa inicia la extensión a partir del extremo 3' de los oligonucleótidos específicos de alelos, con la consiguiente ligación al cebador específico de locus. De esta forma, durante el proceso de extensión y ligación, se generan de forma simultánea un número de moldes igual al de SNPs incluidos en el ensayo y que servirán para la realización de una posterior PCR (*Figura 14A*).

Las condiciones de amplificación son:

- Estabilización a 37 °C durante 10 minutos y desnaturalización a 95 °C durante 3 minutos.

- Amplificación con repetición de 35 ciclos: desnaturalización a 95 °C durante 35 segundos, anillamiento a 56 °C durante 35 segundos y elongación a 72 °C durante 2 minutos.

- Elongación final a 72 °C durante 10 minutos y a 4 °C durante 5 minutos.

Una vez realizada la PCR, se procede a desnaturalizar los fragmentos amplificados y a retener las hebras que portan los fluorocromos. Estas últimas se hibridan con las secuencias complementarias a la secuencia señal, que se localizan en los distintos puntos de genotipado del array.

Una vez realizado el ensayo, la señal emitida por ambos fluorocromos es registrada a través de un sistema de imagen confocal de 0,5 μm que permite medir intensidades de fluorescencia a 532 nm (Cy3) y 635 nm (Cy5), denominado *Beadarray Reader* (Illumina). Este sistema lee las matrices de 96 muestras en ambos canales de color en un breve espacio de tiempo.

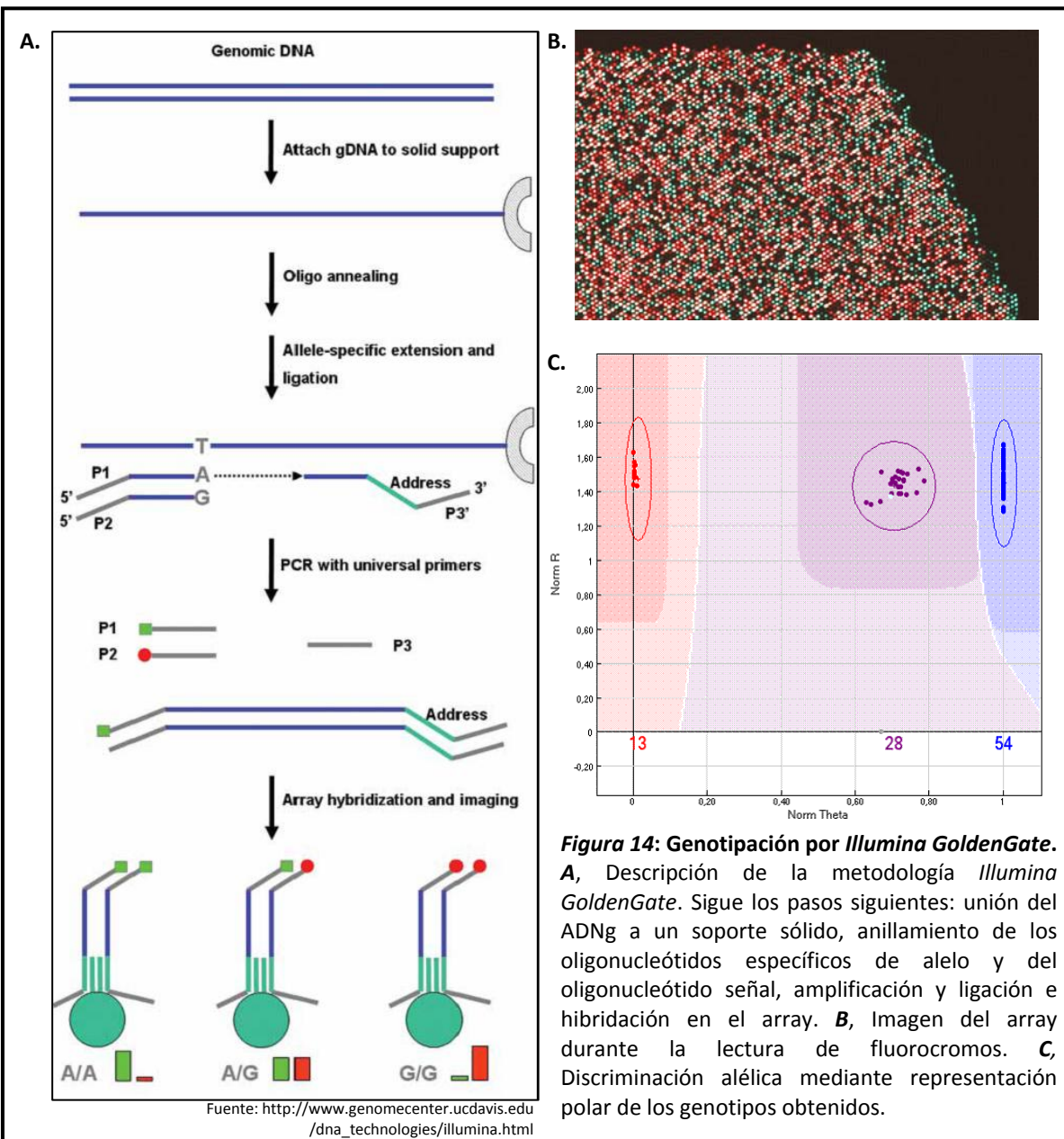
Dependiendo del genotipo de una muestra para un determinado polimorfismo, la razón entre la señal generada por el fluorocromo Cy3 y la generada por el fluorocromo Cy5 será distinta. La imagen que se obtiene del programa es una composición no real de una mezcla entre color verde y rojo, de forma que los puntos de genotipo que aparecen en alguno de estos dos colores representarían genotipos homocigotos, mientras que la aparición de puntos de genotipado amarillos indicaría la existencia de *loci* heterocigotos (*Figura 14B*).

Una vez medidas estas intensidades, se obtiene la representación cartesiana y polar de las intensidades de ambos fluorocromos. El propio sistema de la plataforma permite asignar un determinado genotipo a cada una de las muestras, aunque de forma manual se puede incluir o descartar genotipos marginales (*Figura 14C*).

Para nuestro estudio utilizamos una OPA (*custom oligonucleotide pool*) de 768 SNPs que comprendían 122 genes implicados en cáncer dentro de un proyecto piloto para la puesta en marcha de la plataforma de genotipado *Illumina GoldenGate* en el grupo de Genética Humana, cuyo diseño se presenta en el *Anexo I.9*.

III. MATERIALES Y MÉTODOS

III.3. Genotipado de los SNPs seleccionados



III.4. ENSAYOS FUNCIONALES

III.4.1. Ensayos *in silico*

Para evaluar la importancia funcional de las variantes asociadas a la enfermedad, se utilizaron las herramientas bioinformáticas explicadas anteriormente en el apartado III.2.2.4.

III.4.2. Ensayos *in vitro*: clonaje

Se realizaron el ensayo *in vitro* de clonaje del fragmento que incluye a rs3213180, con el fin de testar la actividad del extremo 3'UTR de *E2F1*.

III.4.2.1. Vectores de clonaje

El vector de expresión de Promega utilizado en el ensayo de clonaje fue pGL3-Control, donde el clonaje se realizó a 3' del gen *Luciferasa* con la enzima de restricción *XbaI* (New England BioLabs) (Figura 15). Esta enzima de restricción, que lineariza el vector, coincide con la utilizada para establecer extremos cohesivos en los insertos y se utiliza siguiendo las instrucciones del fabricante.

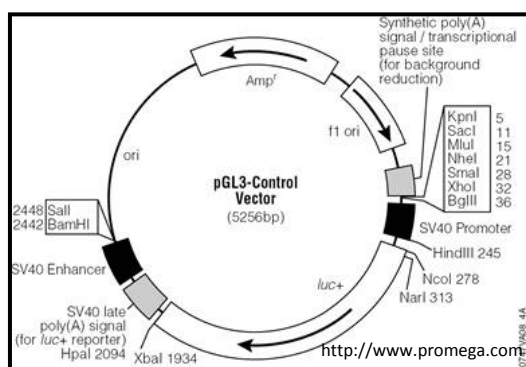


Figura 15: Vector de clonaje pGL3-Control.

Para mantener la linealidad del vector, es necesario tratarlo con SAP. Para ello se utilizan 30 µl de vector digerido que se van a introducir en una mezcla de SAP buffer 10x y 1 U/µl de SAP. Se dispone en un programa de 37 °C durante 15 minutos y 65 °C durante 15 minutos.

III.4.2.2. PCR de clonaje

Una vez seleccionada la secuencia que se quiere amplificar y clonar, y se diseñaron los oligonucleótidos adecuados a los que se les añade una secuencia adyacente y la secuencia diana de la enzima de restricción que permitirá la inserción del fragmento en el vector. Así pues, los oligonucleótidos diseñados para *XbaI* siguieron el modelo que sigue (ver Anexo I.5.):

5'-ATACCGTCTAGA-secuencia del oligonucleótido-3'

Secuencia de apoyo Secuencia diana *XbaI*

En el caso de que la secuencia a clonar fue mayor de 1 Kb y con el fin de evitar introducir errores en la secuencia a clonar, se utilizó una Taq polimerasa de alta fidelidad como es la *PfuTurbo DNA Polymerase* (Stratagene, EE.UU.) en una proporción 1/10 con la ADN polimerasa normal y con un volumen final de 50 µl, para obtener la mayor cantidad de material de clonaje posible.

La purificación del fragmento de clonaje se realizó desde el gel de agarosa, con el fin de cortar la banda de interés de forma específica. Para ello, se utilizó un gel de agarosa ultrapura (Ultrapure TM Agarose, Invitrogen, EE.UU.) en el que se carga todo el volumen de reacción y el kit de extracción de ADN desde gel de agarosa *QIAquick Gel Extraction Kit* (Qiagen, EE.UU.), siguiendo las instrucciones del fabricante.

III.4.2.3. Digestión del inserto

Se crean extremos cohesivos gracias a la digestión con *XbaI*, según las instrucciones del fabricante.

La purificación de los insertos se realiza a través de la extracción desde un gel de agarosa, con el método indicado anteriormente.

III.4.2.4. Ligación de vector e inserto

Para la ligación del vector con el inserto se utilizaron dos métodos diferentes, en función del tamaño del inserto:

a. Rapid DNA Ligation Kit (Roche, Alemania), según las instrucciones del fabricante. Está indicada para insertos de tamaño pequeño y medio.

b. Método clásico de ligación con *T4 DNA ligase* (Invitrogen, EE.UU.), indicada para aquellos insertos mayores de 500 pb. Se calcula una proporción 3:1 de inserto:vector para que la ligación funcione correctamente, teniendo en cuenta el tamaño en pb de cada uno, se siguen las instrucciones del fabricante y se incuba a 4 °C durante la noche.

Para analizar la correcta ligación del vector y el inserto en el sentido adecuado, se llevó a cabo una digestión con *Pst*I (New England BioLabs, Reino Unido), según las instrucciones del fabricante, y se testó en un gel de agarosa.

Posteriormente, a fin de preparar los sistemas de clonaje para la transformación de células competentes, se procedió a la purificación y eliminación de sales por medio del método de los etanoles indicado en el apartado III.3.2.

III.4.2.5. Transformación de células bacterianas competentes

La amplificación del vector con el inserto incluido se realizó por parte de *E. coli* DH5 α electrocompetentes, más efectivas que las bacterias quimiocompetentes.

Obtención de células DH5 α electrocompetentes

- Toma de una muestra de células DH5 α congeladas en glicerol y mantenidas a -80 °C y siembra en una placa de LB¹⁶ sin antibióticos por técnica de estría incubada a 37 °C durante toda la noche.

- Selección de una colonia aislada de la última estría y se crece en un tubo o matraz con 100 ml de LB líquido. Se incuba a 37 °C durante la noche en agitación.

- Se toman 10 ml del cultivo obtenido y se ponen en 1 L de LB (dilución 1/100). Se crecen a 37 °C en agitación y se mide la absorbancia del cultivo cada 2-3 horas hasta que D.O.₆₀₀ = 0,5-0,6, lo que indica que el cultivo ha llegado a la fase estacionaria y todas las células están vivas. Se para el crecimiento al continuar todo el proceso en hielo.

- Centrifugar a 4.000 g durante 20 minutos a 4 °C, se retira el sobrenadante, se resuspenden las células en su volumen original (500 ml aproximadamente) de H₂Od miliQ estéril a 4 °C y se vuelve a centrifugar a 4.000 g durante 20 minutos a 4 °C. A continuación se resuspenden las células en la mitad del volumen original (250 ml aproximadamente) de H₂Od miliQ estéril a 4 °C y se vuelve a centrifugar a 4.000 g

¹⁶ LB: Medio de cultivo Luria-Bertoni que contiene por cada litro de medio 10 g de triptona, 5 g de extracto de levadura y 10 g de NaCl, ajustado a pH 7,5 con NaOH.

durante 20 minutos a 4 °C. Se retira el sobrenadante y se resuspenden las células en 20 ml de H₂O de miliQ estéril a 4 °C con un 10% de glicerol y centrifugar a 4.000 g durante 20 minutos a 4 °C.

- Retirar el sobrenadante y resuspender las células en 1,5 ml de glicerol al 10%, llegando a un volumen final de 3 ml, que se congela a -70 °C en criotubos.

Transformación de células DH5α electrocompetentes

- Se transforman 50 µl de células competentes con 2 µl del producto de clonaje purificado y se disponen en cubetas de electroporación de 0,2 cm (Bio-Rad, EE.UU.). Las condiciones de electroporación son las que siguen: 200 Ω, 2.500 Faraday y 2,5 kV.

- Cultivo de las células electroporadas en 250 µl de medio LB a 37 °C durante la noche en agitación.

- Cultivo de 100 µl de las bacterias transformadas en medio LB selectivo con el antibiótico de resistencia a ampicilina a 100 µg/ml.

III.4.2.6. Extracción del plásmido clonado de bacterias transformadas

La extracción de los vectores portadores de inserto clonados por bacterias competentes se realizó utilizando los métodos de *QIAprep Miniprep Kit* y *QIAprep Midiprep Kit* (Qiagen, EE.UU.), según las instrucciones del fabricante.

III.4.2.7. Mutagénesis dirigida

El ensayo de mutagenesis dirigida según *QuikChange® XL Site-Directed Mutagenesis Kit* (Stratagene, EE.UU.) es un sistema que incluye cambios puntuales en la secuencia de los insertos a estudio, siguiendo las instrucciones del fabricante para su ejecución.

En la reacción de mutagénesis se utiliza la ADN polimerasa *PfuTurbo* para replicar ambas cadenas del plásmido con una alta fidelidad, y dos oligonucleótidos, directo y reverso, que cubren la zona a mutar e incluyen los cambios de interés. Para el diseño de los oligos se utilizó el *QuikChange Primer Design Program* (Stratagene, EE.UU.), que estima los mejores oligonucleótidos para mutar una zona determinada, dejando las bases mutadas en el centro de la secuencia de los oligonucleótidos (*Anexo I.5*). Una vez incorporados los oligonucleótidos al plásmido, se genera una hebra mutante con huecos

sin sellar. La digestión del producto de PCR de mutagénesis con la endonucleasa *DpnI*, específica para la degradación de hebras parentales metiladas o hemimetiladas, permite la selección de los productos mutados. El clonaje del producto y el sellado de los huecos presentes en las cadenas mutantes se llevaron a cabo por células ultracompetentes XL10-Gold tras su transformación siguiendo las instrucciones predeterminadas.

III.4.2.8. Transfección de células eucariotas

Se utilizaron las células HeLa, cultivadas en medio DMEM¹⁷ con un 10% de suero bovino fetal (FBS, *foetal bovine serum*), 100 U/ml de penicilina, 100 µg/ml de estreptomicina y 250 ng/ml de fungizona (anfotericina B).

La transfección de células eucariotas requirió la utilización de *Lipofectamine™ 2000 Transfection Reagent* (Invitrogen, EE.UU.) como vector de transfección, en *Opti-MEM I Reduced Serum Medium*, siguiendo el procedimiento que indica el fabricante. La lipofectamina permite la introducción del plásmido compuesto por el vector y el inserto que se quiere testar, a través de la membrana de la célula eucariota. Por cada pocillo se utilizaron 30 ng del plásmido con *Luciferasa* y 10 ng del vector portador de *Renilla*. Los experimentos de transfección se repitieron cinco veces para obtener unos resultados los más ajustados a la realidad posible. Además, los experimentos fueron acompañados de un control positivo para el sistema de transfección, un vector luciferasa portador del extremo 3'UTR del gen *PTEN*.

III.4.2.9. Ensayo de luciferasa

Las actividades luciferasa y renilla¹⁸, fueron medidas con el *Dual-Luciferase® Reporter Assay System* (Promega, EE.UU.), como indicaba el fabricante. La medida de luciferasa fue normalizada con la de renilla, con el fin de eliminar del análisis las variaciones propias de la técnica.

¹⁷ DMEM: Medio modificado de Dulbecco y Eagle (*Dulbecco's modified Eagle's medium*), es una mezcla de sales enriquecida con aminoácidos y otros componentes esenciales para el crecimiento celular que se utiliza comúnmente como medio de cultivo de células eucariotas. Contiene 4,5 g/L de glucosa y L-glutamina.

¹⁸ Luciferasa es el nombre genérico para las enzimas que producen bioluminiscencia. La más conocida y utilizada es la luciferasa Firefly, de *P. pyralis*, donde la luz es producida por la oxidación de la luciferina a través del ATP, mientras que la reacción de Renilla implica la utilización de iones de calcio.

III.4.2.10. RT-Q-PCR

Se utilizó esta tecnología para la medición de la expresión génica de los sistemas de luciferasa y resistencia a ampicilina de los plásmidos formados por pGL3-Control y los diferentes insertos de 3'UTR de *E2F1*.

En la RT-Q-PCR (*reverse transcriptase quantitative PCR* o PCR cuantitativa o a tiempo real tras una transcripción reversa) los procesos de amplificación y detección se producen de manera simultánea, donde se emplean oligonucleótidos para la amplificación de la región de interés y sondas marcadas con fluorocromos para detectar dicha amplificación a tiempo real. Las sondas de hidrólisis de Roche empleadas en esta técnica son oligonucleótidos que presentan fluorocromos en ambos extremos y tienen una secuencia complementaria aparte del fragmento de ADN que se quiere amplificar. Uno de los fluorocromos actúa como donador de fluorescencia en el extremo 5' y el otro como aceptor de esta fluorescencia en el extremo 3', siguiendo un sistema de detección similar al de las sondas *TaqMan* de discriminación alélica (*Figura 16*). Un detector realiza la medida de esta emisión de fluorescencia de los fluorocromos hidrolizados, que es proporcional a la cantidad de ADN presente, y la representa gráficamente.

La abundancia de un ARNm y el número de moléculas por célula del mismo se correlaciona con el nivel de síntesis de la proteína correspondiente. Mediante la RT-Q-PCR es posible realizar la medición de la cantidad de ARNm de una muestra de forma indirecta tras una reacción de retro-transcripción.

a. Reacción de retro-transcripción (RT-PCR)

La reacción de retro-transcripción de 1 µg de RNA se realizó mediante *M-MLV Reverse Transcriptase* (Invitrogen, EE.UU.), siguiendo las instrucciones del fabricante.

b. Reacción de PCR a tiempo real (Q-PCR):

La reacción de Q-PCR se realizó en una placa de 384 pocillos, donde se preparó previamente una recta patrón con concentraciones crecientes de un ADN complementario (ADNc) control a 1, 1/10, 1/100, 1/1.000, 1/10.000 y 1/100.000 del producido por la reacción de retro-transcripción, con el fin de inferir los datos de amplificación del ADNc de las muestras problema, mediante una recta de regresión lineal simple.

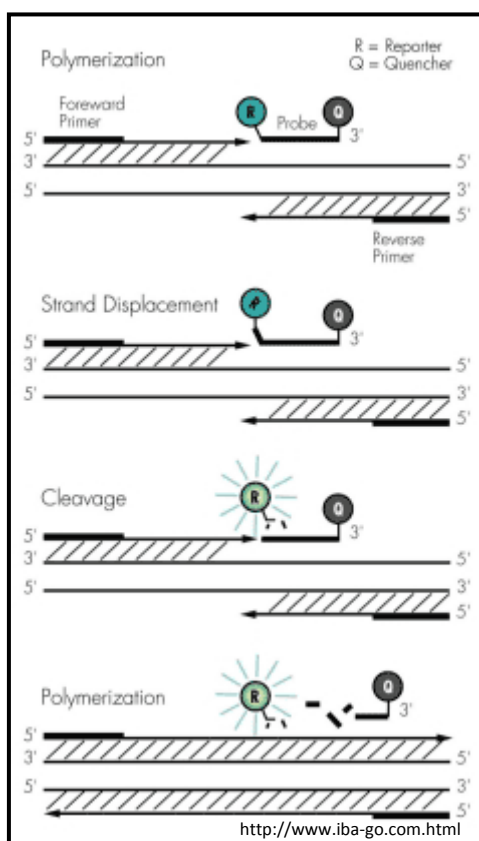


Figura 16: Principio metodológico de la PCR a tiempo real. Se divide en 4 fases: anillamiento de los oligonucleótidos y polimerización, desplazamiento de la sonda, rotura de la sonda y liberación de la fluorescencia.

La mezcla de reacción incluyó los siguientes reactivos:

- Tampón de reacción *TaqMan Universal PCR Master Mix 2x* (UMM), con todos los elementos indicados en el apartado III.3.3.2.
- 2,4 µl cDNA a 1/10 del producido por la reacción de retro-transcripción del ADNc, dispensado mediante el robot *Beckman Coulter FX* en las placas de 384 pocillos cuando fue posible.
- Oligonucleótidos de amplificación y sonda de detección a 10 µM.
- H₂O hasta completar el volumen final de la reacción de 12 µl.

El dispensado de la mezcla fue realizado en la mayoría de las ocasiones con la ayuda del robot *Beckman Coulter FX*, requiriendo un 8% de exceso en previsión de un gasto mayor que un dispensado manual.

La amplificación se llevó a cabo directamente en el *ABI PRISM Sequence Detection System 7900 HT* (Applied Biosystems, EE.UU.), según las siguientes condiciones: desnaturalización de la AmpErasa a 50 °C durante 2 minutos; desnaturalización previa a 95 °C durante 10 minutos; 40 ciclos de desnaturalización a 95 °C durante 15 segundos, y anillamiento y elongación a 60 °C durante 1 minuto.

Cada ensayo fue analizado mediante el software *SDS* v2.1 (Applied Biosystems, EE.UU.), según los datos de fluorescencia obtenidos para las sondas FAM, y se estimó a qué ciclo de Q-PCR ocurre la amplificación del producto de forma exponencial, indicando indirectamente la cantidad de ARNm existente en la muestra original.

Los oligonucleótidos que amplifican las regiones que rodean la zona de unión de la sonda específica para cada gen, así como el número de la sonda de *Universal ProbeLibrary* de Roche correspondiente se indican en el *Anexo I.5.* para los genes *Firefly* y *AmpR*.

III.4.3. Ensayos *in vivo*:

III.4.3.1. Inmunohistoquímica de piezas de tejido mamario embebido en parafina

La técnica de inmunohistoquímica permitió la determinación de la expresión de E2F1 en el tejido mamario embebido en parafina de las muestras indicadas en la *Tabla 3.* Para ello, se siguieron los siguientes pasos:

a. Tinción hematoxilina-eosina:

Permitió estimar el porcentaje de células neoplásicas frente a células normales gracias a la colaboración de anatomopatólogos, si bien al tratarse la mayoría de bloques de tejido normal, el porcentaje de células tumorales fue prácticamente mínimo.

El protocolo de tinción de hematoxilina-eosina se detalla a continuación:

- Incubar durante 1 hora y a 70 °C un corte histológico de 3 µm y desparafinar con xilol por dos veces durante 2 minutos.
- Hidratar con etanol absoluto (2 minutos) y posteriormente con etanol al 95% (2 minutos). Lavar con H₂O durante 2 minutos.
- Teñir con hematoxilina de Mayer durante 15 minutos y lavar con H₂O durante 2 minutos.
- Teñir con eosina durante 2 minutos y eliminar los restos de eosina utilizando etanol al 70%.
- Deshidratar con pasos sucesivos de etanol al 96% (30 segundos), etanol absoluto (2 minutos) y xilol (por dos veces durante 2 minutos).

- Montar el cubre-objetos de la preparación con Eukitt.

b. Inmunohistoquímica:

Las condiciones de unión del anticuerpo al tejido mamario fueron optimizadas para el anticuerpo E2F-1 de UPSTATE (clones KH20 y KH95, EE.UU.). La temperatura, el pH y la dilución se ajustaron con el fin de producir el mejor control.

El proceso automatizado con el equipo *BOND* de *Vision Bio System* incluyó: desparafinación (indicada más arriba), recuperación del antígeno y ensayo de inmunohistoquímica. Para el paso de recuperación del antígeno, se realizó una inducción por calor usando un tampón de citrato 10 mM pH 6, a 96 °C durante 30 minutos. El anticuerpo primario de ratón E2F-1 se añadió al tejido durante al menos 30 minutos a temperatura ambiente en una dilución 1:100, con *Novolink™ peroxidase* como sistema de visualización. Las secciones fueron entonces incubadas con DAB (diaminobenzidina) durante 67 minutos. Posteriormente, las secciones volvieron a ser deshidratadas en alcohol, lavadas en xilol y montadas.

La inmunohistoquímica fue interpretada por un patólogo bajo un sistema ciego de cuantificación de la expresión de E2F1 en los diferentes grupos de muestras. La inmunorreactividad fue considerada positiva cuando las células de los ductos o de los acinos lobulares mostraban un color marrón en los núcleos. El porcentaje de núcleos marcados con E2F1 fue evaluado, independientemente de la intensidad, en las áreas de ductos y acinos lobulares con la mayor expresión del anticuerpo. Las secciones de tejido de mama normal y de amígdala fueron utilizadas como controles positivos externos. Sólo el recuento de los núcleos de ductos fue tenido en cuenta para el análisis estadístico.

III.4.3.2. RT-Q-PCR de piezas de tejido mamario embebido en parafina

Se realizó este ensayo con el ARNm extraído de las piezas de tejido mamario embebido en parafina, utilizando las sondas y los oligonucleótidos de amplificación de las regiones de interés que se indican en el *Anexo I.5.* para los genes *E2F1* y *ACTB* (gen de β -Actina).

III.5. ANÁLISIS ESTADÍSTICOS

En todos los análisis estadísticos referidos, se tomó como valor significativo aquel con un valor de p-valor menor de 0,05.

III.5.1. Estudio de asociación al alelo raro

Para todos los polimorfismos estudiados, se realizó un test exacto de Fisher para verificar que el ensayo no se desviaba del equilibrio de Hardy-Weinberg (HWE *Hardy-Weinberg equilibrium*¹⁹) entre controles, pues aquellos SNPs fuera de HWE en controles fueron desechados del estudio. Asimismo, los OR (*odds ratio*) asociados al alelo menor o raro de cada SNP, sus correspondientes intervalos de confianza al 95% (95% I.C.) y sus p-valores asociados fueron estimados mediante un test chi-cuadrado (χ^2), siempre utilizando como referencia el alelo más frecuente en controles. Para ello se utilizó la herramienta bioinformática *Hardy-Weinberg equilibrium calculator* (<http://ihg2.helmholtz-muenchen.de/cgi-bin/hw/hwa1.pl>).

III.5.2. Estudio de asociación al genotipo

Los riesgos relativos asociados a los genotipos de cada SNP, fueron estimados mediante un test χ^2 y regresión logística no condicionada mediante el programa *SPSS* v15.0 y la herramienta *Hardy-Weinberg equilibrium calculator* antes mencionada. En el análisis de los SNPs, los modelos dominantes, codominantes y recesivos fueron considerados, siempre utilizando como referencia el genotipo de los homocigotos del alelo más frecuente en controles, seleccionando el modelo más adecuado según el test de máxima verosimilitud (codominante *versus* lo demás). Por otra parte, aquellos factores de riesgo conocidos para cáncer de mama como son la edad (establecido el

¹⁹ HWE: Es la situación en la cual la frecuencia de los genotipos para un locus está completamente determinada por la frecuencia de los alelos. Así, si la frecuencia del alelo A es p_A , y la frecuencia del alelo B es p_B , en condiciones de equilibrio de Hardy-Weinberg, la frecuencia de los genotipos se define según: $AA = p_A^2$; $AB = 2p_Ap_B$; $BB = p_B^2$.

corte en la mediana de edad de los controles), la paridad (nulípara frente a unípara y múltipara), la edad de menarquia (establecido el corte en la mediana de edad de menarquia de los controles) y el estado de menopausia (postmenopausia frente a premenopausia) fueron evaluados como potenciales efectos de confusión al incluirlos en un análisis multivariante.

III.5.3. Estudio de asociación al haplotipo y al diplotipo

III.5.3.1. Estimación de haplotipos y diplotipos por *PHASE*

Un haplotipo es un conjunto específico de alelos observados en un único cromosoma. Teniendo en cuenta los genotipos de los SNPs estudiados para cada uno de los genes, procedimos a estimar los posibles haplotipos de éstos. Para ello, se hizo uso del programa *PHASE* v2.0. Este método computacional permite, usando análisis estadísticos bayesianos, realizar tests de permutación para inferir qué haplotipos aparecen en nuestras muestras y en qué frecuencia.

Los archivos obtenidos permiten conocer:

- Los haplotipos estimados dentro del conjunto de muestras del estudio.
- Las frecuencias de los haplotipos en casos frente a controles de población normal, indicando si existe una diferencia significativa (con un p-valor) en la distribución de haplotipos entre ambos grupos.
- El diplotipo de cada caso o control, es decir, los 2 haplotipos más probables para cada una de las muestras en sus 2 cromosomas.
- Estimaciones de los parámetros de recombinación de la región considerada.

III.5.3.2. Análisis estadístico de haplotipos y diplotipos

Utilizando los resultados obtenidos mediante *PHASE* v2.0, se estimó el riesgo relativo asociado a cada haplotipo y a cada diplotipo utilizando un test χ^2 . En el análisis de haplotipos se consideró como referencia el haplotipo más frecuente en controles, y en el de diplotipos, los homocigotos del haplotipo más frecuente en controles. El OR no pudo ser estimado en aquellos haplotipos o diplotipos que aparecían únicamente en uno de los 2 grupos (casos o controles).

III.5.4. Estudio de asociación del genotipo con variables fenotípicas

También se evaluó la asociación entre los polimorfismos genotipados y varias características clínicas y tumorales por regresión logística, con el fin de determinar su implicación en la enfermedad como potenciales efectos modificadores de la expresión fenotípica y su interacción con el genotipo. Para ello, se comparó cada una de las categorías dentro de cada variable evaluada considerando sólo el grupo CME, puesto que aquellos valores comunes para casos y controles ya habían sido evaluados al ser considerados los factores de confusión para el cáncer de mama o bien no tener suficiente representación en la muestra. Las características que se consideraron fueron: metástasis al momento del diagnóstico (según el valor “M” de TNM), histología del tumor (invasivo frente a *in situ*), grado de desarrollo del tumor (grado > 1 frente a grado 1), el tamaño del tumor (> 2 cm frente a ≤ 2 cm, según el valor “T” de TNM), presencia de ganglios linfáticos con metástasis (según el valor “N” de TNM) y estado de los receptores de estrógenos y progesterona (los que tienen una mayor representación en nuestra población). La estratificación de la muestra en función de las diferentes categorías establecidas y la comparación con el grupo CoCM sólo se llevó a cabo cuando el análisis anterior llevaba a diferencias entre grupos según el genotipo, a fin de determinar su implicación en el desarrollo de la enfermedad.

Asimismo, se establecieron análisis de supervivencia, tanto en función de las distintas variables personales, tumorales o clínicas, como en relación a genotipos determinados, utilizando un test de Kaplan-Meier y por regresión de Cox.

III.5.5. Estudio estadístico de los ensayos funcionales

III.5.5.1. Análisis estadístico de transfección

Los resultados de los experimentos de transfección en células HeLa se expresaron con medias y desviaciones estándar de resultados de expresión de luciferasa normalizados con los datos de expresión de renilla, en experimentos repetidos al menos cinco veces en ensayos independientes, con el fin de llegar a resultados estadísticos

ajustados. Las diferencias entre la actividad luciferasa de los diferentes productos de clonaje estudiados se analizaron mediante un test t de Student.

III.5.5.2. Análisis estadístico de inmunohistoquímica y RT-Q-PCR

Los resultados derivados de inmunohistoquímica y RT-Q-PCR se analizaron mediante un test ANOVA y regresión logística lineal.

IV. RESULTADOS

IV.1. GENERALIDADES DE LOS DATOS CLÍNICOS

IV.1.1. Definición de los factores clínicos de confusión

Se realizaron análisis estadísticos por el test de Fisher con el fin de determinar si existían diferencias entre CoCM y CME en aquellas características clínicas comunes a ambos grupos descritas previamente en la *Tabla 2*. Con ello se pretendió definir los factores de confusión para el cáncer de mama en nuestro grupo de muestras, que se utilizarían en la corrección de los análisis estadísticos de asociación del genotipo a la enfermedad.

Las características clínicas que se observaron se distribuían de una manera estadísticamente diferente entre casos y controles para cáncer de mama, y fueron las que se señalan a continuación: edad de diagnóstico para CME o edad en el momento de participación en el estudio para CoCM ($p=1,08 \times 10^{-6}$), paridad o número de hijos nacidos vivos ($p=5,53 \times 10^{-4}$), estado de menopausia ($p=5,27 \times 10^{-10}$), historia clínica de cáncer genérico ($p=0,001$) e historia clínica de cáncer de mama anterior ($p=0,036$). Sin embargo, sólo se consideraron aquellas características con una representación en la población de CoCM y CME suficiente, esto es, que no tenían más de un 50% de datos desconocidos en alguno de los dos grupos de muestras. Esta limitación nos llevó a definir las características de edad, paridad y estado de menopausia, como aquellas intrínsecamente diferentes en nuestra población, ya ampliamente descritas como factores de confusión para cáncer de mama en trabajos previos. A estas tres características se le sumó la edad de menarquia como un factor de confusión que también se ha tenido en cuenta en los análisis estadísticos del presente trabajo. Si bien la edad de menarquia no es estadísticamente diferente entre casos y controles en nuestra colección, sí que se ha descrito como un factor de consideración en el cáncer de mama, junto a los ya mencionados (Cuzick, 2008).

Una posibilidad que debe ser tomada en cuenta para las asociaciones detectadas en la presente tesis doctoral, es que éstas sean consecuencia de factores de confusión no valorados, como puede ser el grupo étnico de las muestras estudiadas. Sin embargo, la

estratificación de la población fue descartada tras realizar un análisis de 28 marcadores moleculares bialélicos localizados a lo largo de todo el genoma y no asociados a ninguna patología en un subgrupo de 343 muestras estudiadas en este trabajo tomadas al azar (163 casos y 180 controles) (Milne *et al.*, 2006).

IV.1.2. Supervivencia en función de las características clínicas

El ensayo de supervivencia para las diferentes características clínicas se realizó con el fin de definir la importancia de cada una de ellas en la evolución de la enfermedad. Para ello se consideraron sólo aquellas características para las que hubo un número mínimo de datos entre pacientes vivas y fallecidas, al ser éstas últimas un grupo muy reducido de tan sólo 28 pacientes, lo que limitaba el análisis estadístico de los datos. En total se analizaron 338 pacientes con datos del estado vital.

Mediante el análisis de Kaplan-Meier y la regresión de Cox, se estableció que las pacientes con una peor supervivencia a la enfermedad eran aquellas que presentaban un mayor grado de tumor (OR=3,97, 95% I.C. 1,89-8,36, $p=2,86 \times 10^{-4}$), un mayor tamaño del tumor (OR=3,98, 95% I.C. 1,80-8,77, $p=0,001$), presencia de metástasis en los ganglios linfáticos axilares anexos a la mama en el momento del diagnóstico (OR=17,53, 95% I.C. 5,27-58,29, $p=2,98 \times 10^{-6}$) y presencia de metástasis general a otros tejidos diferentes de los ganglios linfáticos axilares en el momento del diagnóstico (OR=21,08, 95% I.C. 8,14-54,57, $p=3,37 \times 10^{-10}$). Sin embargo, se pudo determinar que factores como la historia clínica del paciente o el estado de los marcadores tumorales evaluados no influyen en la supervivencia de las pacientes incluidas en nuestro estudio. Los datos correspondientes a los análisis estadísticos evaluados, así como la representación gráfica de las diferencias de supervivencia de aquellas características con un p-valor $<0,05$ se presentan en la *Tabla 5* y la *Figura 17*, respectivamente.

Tabla 5: Análisis de supervivencia en función de características clínicas y tumorales por regresión de Cox.

Característica clínica y tumoral	N=338		
	OR	I.C. 95%	p-valor
Historia clínica de cáncer de mama anterior	0,48	0,06-3,53	0,468
Historia familiar de cáncer de mama	2,07	0,93-4,62	0,075
Histología del tumor	3,93	0,53-29,10	0,181
Grado del tumor	3,97	1,89-8,36	2,86x10⁻⁴
Tamaño del tumor	3,98	1,80-8,77	0,001
Metástasis de ganglios linfáticos al diagnóstico	17,53	5,27-58,29	2,98x10⁻⁶
Metastasis general al momento del diagnóstico	21,08	8,14-54,57	3,37x10⁻¹⁰
ER	0,55	0,23-1,31	0,174
PR	0,47	0,20-1,09	0,080

Los p-valores estadísticamente significativos están indicados en negrita roja. Los resultados marginales se indican en negrita negro.

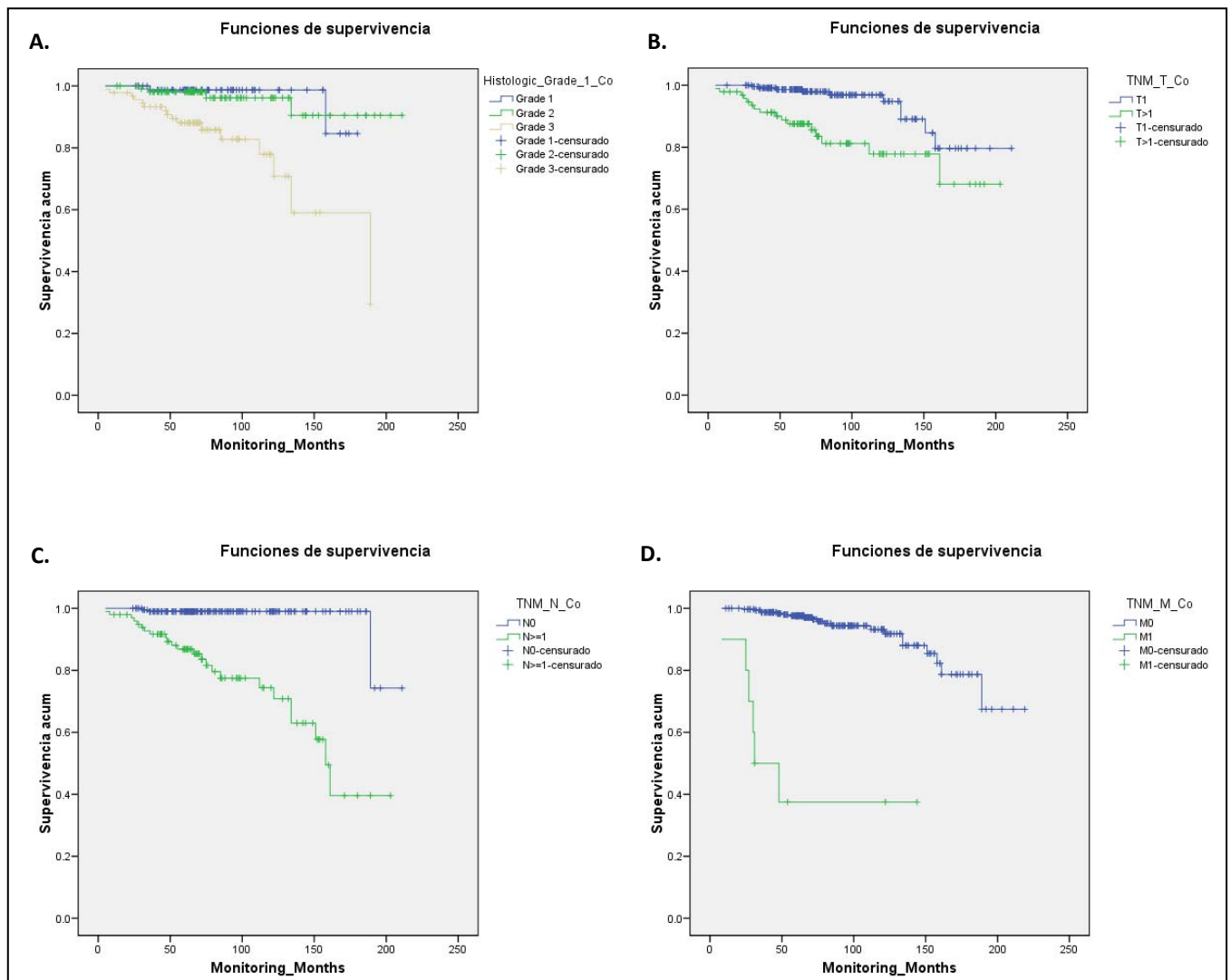


Figura 17: Representación gráfica de los resultados de supervivencia significativamente diferentes en función de las características clínicas según el análisis de Kaplan-Meier. A, Grado del tumor; B, Tamaño del tumor; C, Metástasis a ganglios linfáticos en el momento del diagnóstico; D, Metástasis general en el momento del diagnóstico.

IV.2. ESTUDIO DE ASOCIACIÓN A CÁNCER DE MAMA ESPORÁDICO DE GENES IMPLICADOS EN LA REPARACIÓN DEL ADN

IV.2.1. Genes y SNPs seleccionados de la ruta de reparación del ADN

Para el presente trabajo se tomó como punto de referencia en el estudio de susceptibilidad a padecer cáncer de mama al grupo de genes cuyas proteínas están implicadas en la protección frente al daño que puede ocurrir en el ADN, reparación de dicho daño y regulación de la entrada en apoptosis cuando el daño es tan profundo que la reparación no es viable. Se realizó esta primera aproximación debido a que los genes conocidos hasta el momento del comienzo del proyecto cuyas mutaciones derivaban en cáncer de mama familiar eran *BRCA1*, *BRCA2*, *CHEK2* y *ATM* (Oldenburg *et al.*, 2007), entre otros, y las proteínas que codificaban intervenían, fundamentalmente, en las funciones arriba descritas.

Los genes seleccionados de la ruta de reparación del ADN fueron 13 miembros de la familia de genes *FANC*: *FANCA*, *FANCB*, *FANCC*, *FANCD1/BRCA2*, *FANCD2*, *FANCE*, *FANCF*, *FANCG*, *FANCI*, *FANCL/BRIP1*, *FANCL*, *FANCM*, *FANCN/PALB1*, y 6 genes relacionados con los anteriores: *ATM*, *ATR*, *BRCA1*, *H2AX*, *OGG1* y *USP1* (ver *Figura 4* para más detalle). Sus funciones específicas y las referencias bibliográficas que llevaron a su elección se presentan en el *Anexo I.10*.

La utilización de las herramientas bioinformáticas y el seguimiento de los criterios de selección de SNPs descritos en el apartado III.2.2.1, llevó a la selección de 104 SNPs. En el *Anexo I.11* se presentan sus características más relevantes, como son: la posición cromosómica, la posición en el gen (que podría indicar su función putativa en la expresión génica), el cambio nucleotídico que produce, el cambio de aminoácido que implica en el caso de ser un SNP codificante, y la MAF en una población caucásica estudiada por el proyecto HapMap.

IV.2.2. Relación de *FANCD2* con el cáncer de mama esporádico

En este apartado, se describe con detalle los resultados obtenidos en el estudio de asociación del gen *FANCD2* con el cáncer de mama esporádico.

IV.2.2.1. Estudio de asociación de *FANCD2*

El análisis de frecuencia alélica de los dos SNPs evaluados inicialmente en el gen *FANCD2*, rs722509, un cambio localizado en el intrón 19 y rs2272125, que lleva a un cambio sinónimo p.Leu1366Leu en el exón 42 del gen (ver *Figura 18*), reveló que el SNP rs2272125 presenta una MAF mayor en casos de cáncer de mama esporádico comparada con controles (0,13 *versus* 0,10, respectivamente), lo que supone una diferencia estadísticamente significativa ($p=0,005$). Por otra parte, la comparación de las frecuencias de genotipos estimó que el riesgo para el desarrollo de cáncer de mama esporádico asociado a rs2272125, siguiendo un modelo codominante, es más del doble en homocigotos raros “CC” (OR=2,77, 95% I.C. 1,08-7,13, $p=0,030$) que en heterocigotos “AC” (OR=1,28, 95% I.C. 1,02-1,62, $p=0,040$), comparándolo con los homocigotos normales “AA”. El modelo que mejor se aproximaba era un modelo codominante de parámetro simple, con un riesgo relativo por alelo de 1,36 (95% I.C. 1,09-1,68, $p=0,006$).

Esta asociación estadísticamente significativa se pierde después de ajustar por los factores de confusión comentados previamente: edad de diagnóstico o de recogida de muestra, edad de menarquia, número de hijos y estado de menopausia (OR=1,27, 95% I.C. 0,90-1,79, $p=0,172$), aunque el sentido del riesgo asociado se mantuvo en los mismos niveles. No se observó ninguna asociación entre la enfermedad y el otro SNP analizado en *FANCD2*, rs722509. Los resultados de los análisis de genotipos se representan en la *Figura 19*.

IV. RESULTADOS

IV.2. Estudio de asociación a cáncer de mama esporádico de genes implicados en la reparación del ADN

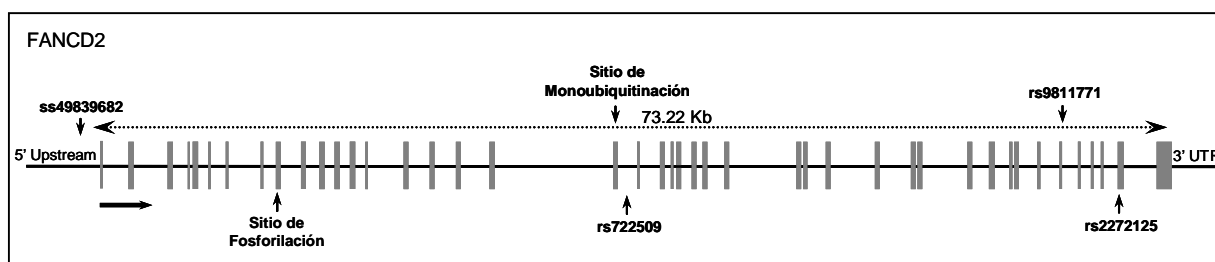


Figura 18: Representación esquemática del gen *FANCD2*. Los exones se indican como barras grises verticales. Se indica la localización de los SNPs genotipados inicialmente, rs722509 y rs2272125, así como aquellos estudiados en una genotipación posterior, ss49839682 (rs33917318) y rs9811771. Asimismo, se indica el tamaño del gen en Kb y la localización de los motivos funcionales de la proteína *FANCD2*: sitio de fosforilación y sitio de monoubiquitinación.

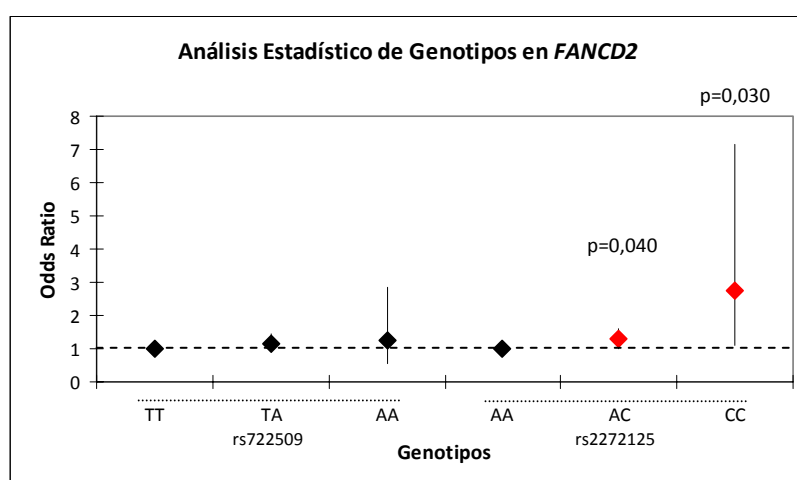


Figura 19: Representación gráfica del análisis estadístico de genotipos en *FANCD2*. Se indican los ORs (rombos, en rojo con estadística significativa) e intervalos de confianza (líneas verticales). Los p-valoros significativos están indicados.

Los resultados derivados de la comparación de haplotipos y diplotipos estimados para *FANCD2* entre casos y controles se presentan en la *Figura 20A* y la *Figura 20B*, respectivamente.

El análisis mediante la herramienta PHASE v2.0 desveló la existencia de un haplotipo común, "TA" o "H1", encontrado en ambos grupos, CME y CoCM, y que lleva los alelos comunes "T" de rs722509 y "A" de rs2272125. Además, estimó los diplotipos para cada muestra, así como sus frecuencias en cada grupo, y se estableció como referencia el diplotipo compuesto por el haplotipo más frecuente en controles, "TA/TA" (portador de los dos cromosomas más comunes).

La comparación de frecuencias de haplotipos entre casos y controles para el haplotipo "TC" o "H2", que lleva el alelo raro "C" de rs2272125, muestra un OR estimado con respecto a "TA" de 4,46 (95% I.C. 1,51-13,21, $p=0,007$), lo cual sigue indicando una mayor presencia del alelo raro "C" del SNP codificante sinónimo entre los casos con

cáncer de mama esporádico, como ocurría con los análisis de frecuencias alélicas y genotípicas (ver *Figura 20A*).

Asimismo, el análisis de los diplotipos determinó que llevar el cromosoma de riesgo “TC” en el diplotipo “TA/TC” estaba asociado a un incremento en el riesgo para el cáncer de mama de 3,57 (95% I.C. 1,17-10,90, $p=0,030$), relativo al diplotipo de referencia “TA/TA”. Más aún, todas aquellas mujeres que se encontraron, portaban el haplotipo “TC” en otras combinaciones que no fueran dentro de un diplotipo “TA/TC”, lo llevaban siempre en combinación con el haplotipo “AC”, formado por los alelos raros de ambos SNPs, y de manera exclusiva entre las pacientes con cáncer de mama esporádico. Debido a la imposibilidad del análisis estadístico del diplotipo “TC/AC”, al no estar presente en el grupo de referencia de controles para cáncer, se combinaron todos los portadores de “TC”, y la comparación de sus frecuencias respecto a “TA/TA” confirmó que su presencia y, en consecuencia, la presencia del alelo menor “C” de rs2272125, incrementaba el riesgo a un OR de 4,59 (95% I.C. 1,54-13,63, $p=0,006$) (*Figura 20B*).

IV.2.2.2. Búsqueda de la variante causal de la asociación de *FANCD2*

A pesar del riesgo para cáncer de mama esporádico observado asociado al SNP rs2272125, es improbable que éste sea el agente causal del mismo debido a su carácter sinónimo. Además, los análisis *in silico* realizados utilizando la herramienta PupaSuite proponen que este SNP no se encuentra en ninguna región activadora o silenciadora del “*splicing*” (ESE, *exonic splicing enhancer* y ESS, *exonic splicing silencer*, respectivamente), localizada en el exón 42 donde se encuentra. Descartada, en principio, la opción de que rs2272125 fuera el cambio causal del riesgo observado, se postuló que dicho efecto podría ser debido a una o varias variantes que se encuentran en el mismo bloque de DL que rs2272125.

IV. RESULTADOS

IV.2. Estudio de asociación a cáncer de mama esporádico de genes implicados en la reparación del ADN

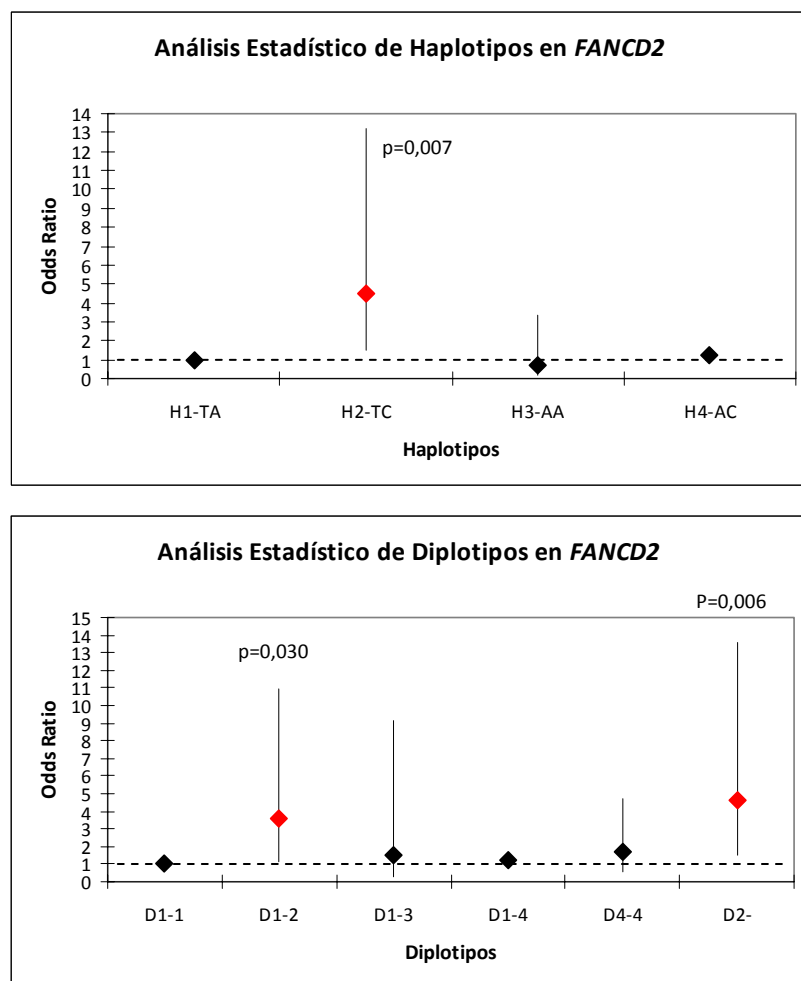


Figura 20: Representación gráfica del análisis estadístico de haplotipos (A) y diplotipos (B) en *FANCD2*. Se indican los ORs (rombos, en rojo con estadística significativa) e intervalos de confianza (líneas verticales). Los p-valoros significativos están indicados. Los haplotipos detectados son: "TA", "TC", "AA" y "AC". Los diplotipos se han nombrado según: D1-1, "TA/TA"; D1-2, "TA/TC"; D1-3, "TA/AA"; D1-4, "TA/AC"; D4-4, "AC/AC"; D2-, cualquier portador del haplotipo "TC".

Se analizó dicho bloque de DL utilizando las bases de datos públicas mencionadas en los apartados III.2.2.2. y III.2.2.3., que aplicaban dos métodos diferentes de estimación de bloques de DL, *Confidence Interval Method* y *4 Gamete Rule*, y se estableció que dicho bloque se extiende a lo largo de todo el gen *FANCD2*, cubriendo más de 100 Kb (ver *Figura 21*). Además de *FANCD2*, según el método de análisis tomado en cuenta, se incluían diferentes genes, la mayoría sin función conocida o descritos solamente a nivel bioinformático. El gen supresor de tumor *VHL*, implicado en diversos carcinomas (Cowey and Rathmell, 2009; Meyer-Rochow *et al.*, 2008) fue la excepción a la falta de genes con funcionalidad patente. Sin embargo, su lejanía de *FANCD2* y la falta de consenso en su inclusión en el mismo bloque de DL que *FANCD2*, hizo rechazar su intervención en la patología como portador de los elementos causales del riesgo observado. Así pues, tomadas en cuenta las consideraciones anteriormente descritas, se retomó la hipótesis de *FANCD2* como el gen implicado en el cáncer de mama esporádico y portador de los elementos causales de dicho riesgo.

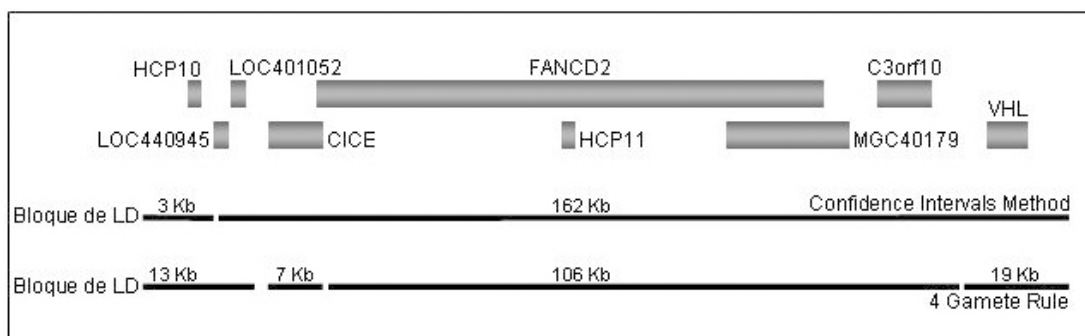


Figura 21: Bloques de DL de la región que engloba a FANCD2 y a los genes que lo rodean. Las cajas grises representan a los genes. Las líneas sólidas representan los bloques de DL obtenidos por los métodos *Confidence Intervals Method* y *4 Gamete Rule*.

Las variantes causales se buscaron entre los portadores del haplotipo de riesgo “TC” (aquellas pacientes “TA/TC” y “TC/AC”). Se procedió, por tanto, al estudio de las regiones descritas como funcionales dentro de la proteína FANCD2 por secuenciación; éstas son: el motivo de fosforilación, localizado en el exón 9, y el motivo de monoubiquitinación, localizado en el exón 19 (ver *Figura 18* para mayor detalle). Sin embargo, no se encontró ningún cambio en las regiones señaladas con respecto a la secuencia de referencia (NC_000004).

Se trató, pues, de caracterizar la región que rodea a rs2272125 (del exón 38 al exón 42), así como la zona promotora (1.500 pb antes del codón de inicio) y 3’UTR, estas dos últimas como dominios con funcionalidad putativa, a través de la tecnología *dHPLC*. Se encontraron tres variantes en la región promotora: rs781871 (sin MAF descrita previamente), rs17032268 (NCBI MAF=0,22) y el SNP ss49839682 detectado por primera vez en nuestro grupo (posteriormente nombrado como rs33917318, a -544 pb del codón de inicio). A 13 pb corriente abajo del exón 38, en el intrón 38, se localizó el cambio rs9811771, con MAF desconocida. Finalmente, se detectaron cuatro SNPs en la zona 3’UTR, estando uno de ellos, rs7647987, localizado antes de la secuencia de poliA (NCBI MAF=0,12), mientras que los otros tres se encontraban después del poliA: rs3826, rs11716842, y un polimorfismo detectado por primera vez, ss49839681 (posteriormente nombrado como rs34305518, a +421 pb del codón de parada), todos con MAF desconocida (ver *Tabla 6* para más detalle). Para ambos SNPs detectados *de novo* aportamos, además, frecuencias alélicas en población española que se han cargado en la base de datos del NCBI.

IV. RESULTADOS

IV.2. Estudio de asociación a cáncer de mama esporádico de genes implicados en la reparación del ADN

Tabla 6: Variantes encontradas por dHPLC y secuenciadas para confirmación, en muestras que portan el haplotipo “TC”.

Muestra	Promotor			Intrón 38	Exón 43			
	rs781871	rs33917318	rs17032268	rs9811771	rs7647987	rs34305518	rs3826	rs11716842
CME-1	T/T	T/C	C/T	A/A	A/A	-	-	-
CME-2	T/A	T/T	C/C	A/A	A/A	-	-	-
CME-3	T/T	T/C	C/T	A/G	A/G	T/A	T/G	C/T
CME-4	T/A	T/T	C/C	A/A	A/A	T/A	T/T	C/T
CME-5	T/T	T/T	C/C	A/A	A/A	T/T	T/T	C/C
CME-6	-	T/T	C/C	A/A	A/A	T/T	T/T	C/C
CME-7	T/T	T/T	C/C	A/G	A/A	T/T	T/T	C/C
CME-8	T/T	T/T	C/C	A/A	A/A	-	-	-
CME-9	-	-	-	A/A	A/A	-	-	-
CME-10	-	T/T	C/C	A/A	A/A	T/T	T/T	C/C
CME-11	-	-	-	A/A	A/A	-	-	-
CME-12	T/A	T/T	C/C	A/A	A/G	T/T	T/T	C/C
CME-13	T/T	T/C	C/T	A/G	A/G	T/T	T/T	C/C
CME-14	T/T	T/T	C/C	A/G	A/G	T/A	T/G	C/C
CoCM-1	T/T	T/C	C/C	A/A	A/G	T/T	T/T	C/C
CoCM-2	T/T	T/C	C/T	A/G	A/G	-	-	-
CoCM-3	T/T	T/T	C/C	A/A	A/A	T/A	T/T	C/C

Las muestras que comparten secciones del mismo haplotipo 2 variante (“H2V”) están sombreadas en gris. Las variantes localizadas entre las líneas discontinuas fueron seleccionadas para un estudio más exhaustivo.

Todas las variantes detectadas en los portadores del haplotipo 2 (“H2”) “TC” estaban en alto DL y, de acuerdo con el programa PHASE v2.0, se encontraban encuadradas dentro de un mismo haplotipo, denominado haplotipo 2 variante “H2V” (CTGGAG, zona sombreada en gris en la *Tabla 6*). Cuatro personas portaban secciones del “H2V” y una quinta el haplotipo entero, representando todas a un 30% de los portadores de “TC”.

Posteriormente se genotiparon cuatro de las ocho variantes mencionadas en 190 personas no portadoras de “TC” (95 CME y 95 CoCM) para poder determinar si alguna de ellas se acumulaba en portadores de “TC”. Éstas fueron: rs33917318, rs17032268, rs9811771 y rs7647987, elegidas según su posible función dentro del gen y porque comparten la mayor parte de “H2V”. Sólo dos de las cuatro variantes seleccionadas (rs33917318 y rs9811771) parecían estar presentes de modo más frecuente en portadores de “TC” que en no portadores, aunque ninguna con un resultado estadísticamente significativo (OR=2,04, 95% I.C. 0,73-5,66, p=0,200 y OR=1,68, 95% I.C. 0,61-4,62, p=0,300, respectivamente). En un genotipado adicional de 1.100 muestras no portadoras del haplotipo “TC” para estos dos SNPs, no se observó ningún cambio en la significación, aunque la tendencia a la acumulación del alelo menor en los portadores de “TC” se mantuvo (ver *Tabla 7* para más detalle).

Tabla 7: Frecuencias alélicas observadas para cuatro SNPs putativamente funcionales en portadores del haplotipo "TC".

SNP	Portadores de "TC"		190 No-portadores de "TC"		1.100 No-portadores de "TC"	
	N _{am} /N _a	MAF	N _{am} /N _a	MAF	N _{am} /N _a	MAF
rs33917318	5/30	0,17*	34/380	0,09*	248/2166	0,11*
rs17032268	4/26	0,13	34/346	0,10	-	-
rs9811771	5/34	0,15**	35/376	0,09**	265/2154	0,12**
rs7647987	6/28	0,18	54/298	0,18	-	-

N_{am}, número de alelos menores; N_a, número total de alelos; MAF, frecuencia del alelo menor. * p (comparación de MAFs) = 0,2 y 0,4, usando 190 y 1.083 no-portadores de "TC", respectivamente. ** p (comparación de MAFs) = 0,3 y 0,7, usando 188 y 1.077 no-portadores de "TC", respectivamente.

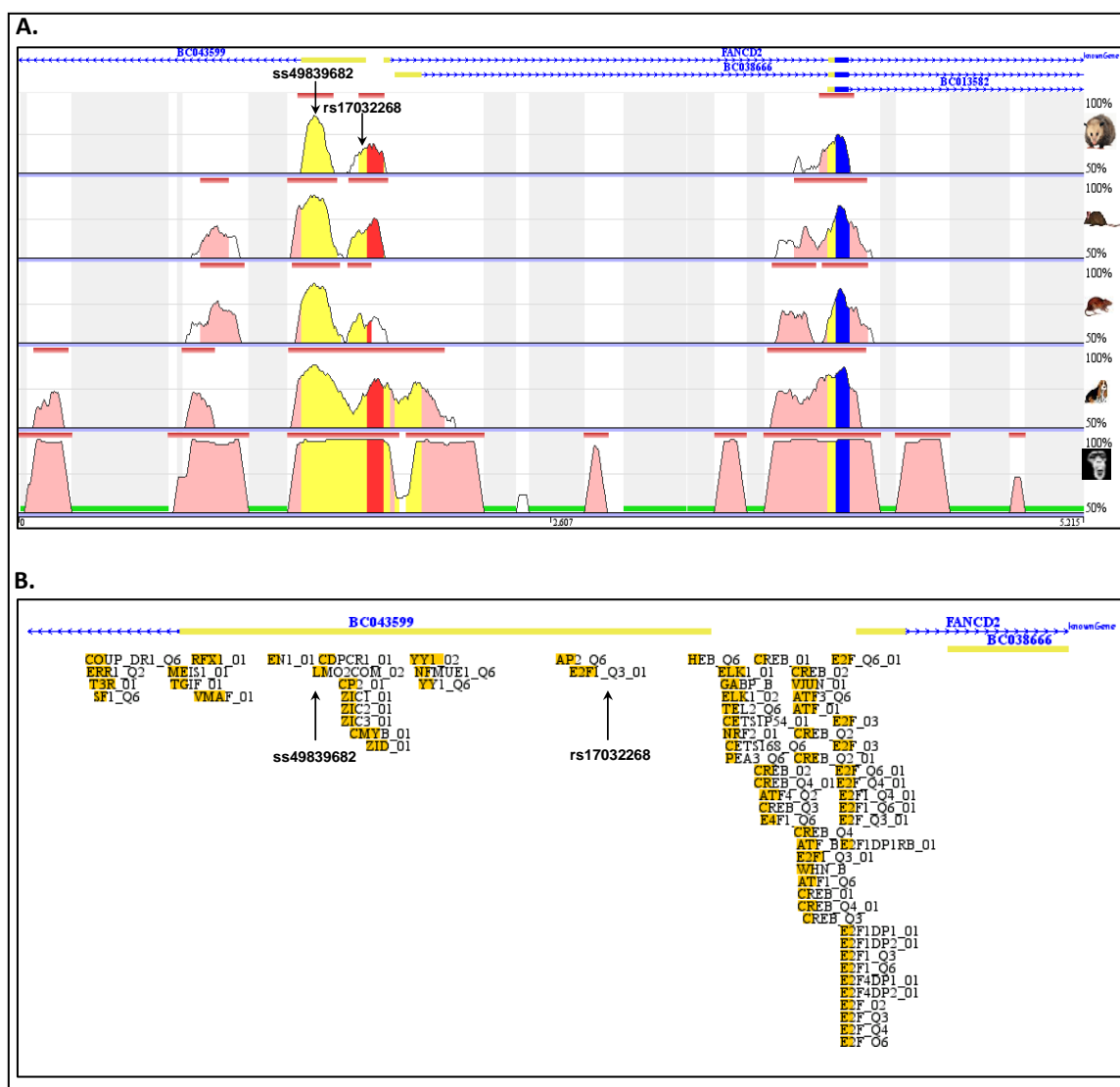


Figura 22: **A**, Representación del estudio filogenético en mamíferos para el gen *FANCD2* según *ECR Browser*. Las cajas azules indican los exones. La identidad de secuencia se valora entre 50-100% en las diferentes especies (margen derecho), indicada en azul cuando son regiones transcritas o en rosa cuando no se transcriben. **B**, Representación de la localización de dianas de factores de transcripción según *ECR Browser* estimadas *in silico*. Los SNPs ss49839682 (rs33917318) y rs17032268, localizados en el extremo 5' de *FANCD2*, están señalados en ambas figuras.

IV.2.2.3. Análisis *in silico* de las variantes putativas causales

Se realizó un análisis de conservación de secuencia de las cuatro variantes que conforman “H2V” en *FANCD2* en una serie de especies de mamíferos: *H. sapiens* versus *P. troglodytes*, *C. familiaris*, *M. musculus*, *R. norvegicus* y *M. domestica*. Sólo los SNPs rs33917318 y rs17032268 (y sus regiones flanqueantes) se mantienen filogenéticamente, como se indica en las *Figura 22A* y *Figura 22B*, lo cual supone una posible funcionalidad en la expresión génica de *FANCD2*.

Un análisis más exhaustivo de la región conservada en la que se encuentran los SNPs arriba mencionados, determinó que estas dos variantes se encuentran localizadas en la zona promotora del gen, en regiones con motivos consenso para factores de transcripción (*Figura 23A* y *Figura 23B*, para cada cambio, respectivamente), siendo rs17032268 la más directamente relacionada con estos elementos al estar localizada en el sitio de unión de E2F1. Tanto la variante rs9811771, localizada cerca del sitio de splicing exón/intrón, como rs7647987, cerca de la zona 3'UTR, no se conservan en ninguna especie de mamífero (*Figura 24A* y *Figura 24B*, para cada cambio, respectivamente).

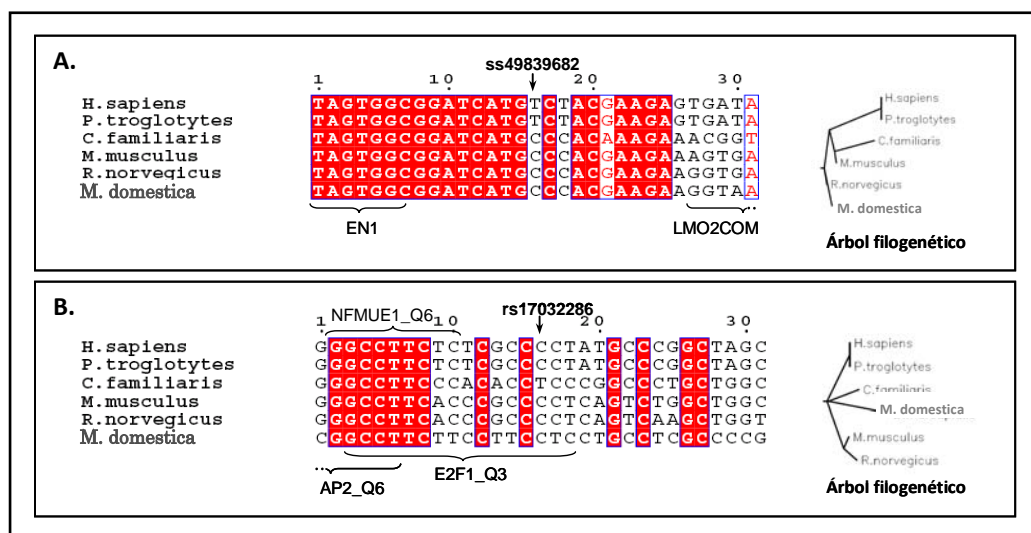


Figura 23. Análisis filogenético de la conservación de secuencia en mamíferos de ss49839682 (rs33917318) (A) y rs17032268 (B), según T-coffee. Las cajas rojas representan las secuencias altamente conservadas y se indican los sitios consenso de unión de factores de transcripción presentes en la región. Se presentan también los árboles filogenéticos y las medidas de divergencia para las regiones de dichos SNPs.

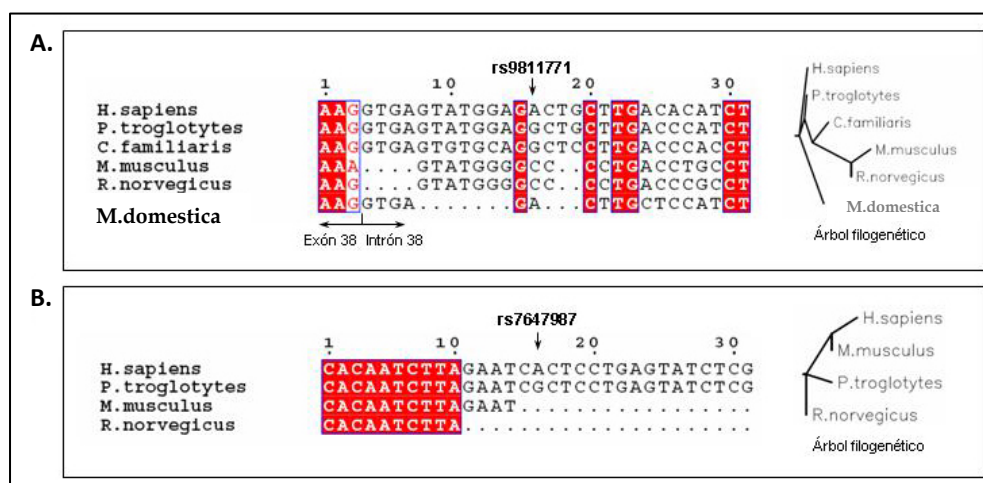


Figura 24. Análisis filogenético de la conservación de secuencia en mamíferos de rs9811771 (A) y rs7647987 (B), según T-coffee. Las cajas rojas representan las secuencias altamente conservadas y se indican los sitios consenso de unión de factores de transcripción presentes en la región. Se presentan también los árboles filogenéticos y las medidas de divergencia para las regiones de dichos SNPs.

IV.2.3. Estudio de asociación de la familia de genes *FANC*

Como consecuencia del riesgo observado para cáncer de mama esporádico de un SNP en *FANCD2*, se consideró oportuno el estudio de la familia de genes *FANC* al completo, llevando a cabo el estudio de un total de 13 genes, incluyendo a *FANCD2*, según: *FANCA*, *FANCB*, *FANCC*, *FANCD1/BRCA2*, *FANCD2*, *FANCE*, *FANCF*, *FANCG*, *FANCI*, *FANCI/BRIP1*, *FANCL*, *FANCM*, *FANCN/PALB1*.

IV.2.3.1. Estudio de asociación de genotipado de los genes *FANC*

El análisis de la distribución de los tres tipos de genotipos (homocigoto normal, heterocigoto y homocigoto raro) de los genes *FANC* entre CME y CoCM permitió identificar un total de 6 SNPs que se desviaban de manera estadísticamente significativa de HWE ($p < 0,05$) en CoCM, descartados en el estudio estadístico posterior. Sólo uno de ellos presentó esta desviación también en los CME. Los SNPs que presentaron una desviación de HWE sin significación estadística se consideraron en los análisis posteriores (detalle en la *Tabla 8*).

IV. RESULTADOS

IV.2. Estudio de asociación a cáncer de mama esporádico de genes implicados en la reparación del ADN

Tabla 8: Análisis estadístico de HWE, indicación de MAF en cada grupo de muestras y análisis estadístico de la distribución de frecuencias alélicas entre CoCM y CME para los genes *FANC*.

Gen	SNP	CoCM		CME		OR (I.C. 95%)	p-valor
		p-valor HWE	MAF	p-valor HWE	MAF		
<i>FANCA</i>	rs2376879	0,120	0,28	1,000	0,26	0,90 (0,74-1,08)	0,237
	rs7190823	0,344	0,33	0,626	0,31	0,91 (0,76-1,08)	0,269
	rs1057042	1,000	0,27	0,441	0,25	0,92 (0,77-1,11)	0,390
	rs2239359	0,196	0,31	0,286	0,33	1,08 (0,90-1,29)	0,397
	rs3785275	0,741	0,25	0,495	0,24	0,93 (0,77-1,13)	0,468
	rs7195066	0,414	0,30	0,836	0,29	0,98 (0,82-1,17)	0,806
	rs2238526	0,670	0,43	0,552	0,44	1,02 (0,87-1,21)	0,780
	rs1800359	0,675	0,49	0,676	0,49	0,99 (0,84-1,17)	0,901
<i>FANCB</i>	rs5935816	0,129	0,39	0,856	0,39	0,98 (0,83-1,17)	0,844
	rs2905223	0,281	0,39	1,000	0,38	0,98 (0,83-1,17)	0,821
<i>FANCC</i>	rs554879	0,857	0,45	0,606	0,45	0,99 (0,83-1,18)	0,917
	rs4647554	0,344	0,39	0,857	0,39	1,01 (0,85-1,21)	0,877
	rs1045276	0,122	0,20	0,676	0,19	0,94 (0,76-1,17)	0,601
<i>FANCD1/BRCA2</i>	rs206116	0,923	0,36	1,000	0,41	1,19 (1,00-1,42)	0,055
	rs3092989	0,096	0,17	0,223	0,15	0,92 (0,73-1,15)	0,447
	rs1801439	0,487	0,03	1,000	0,03	1,02 (0,65-1,60)	0,921
	rs1801406	0,357	0,29	0,471	0,28	0,98 (0,82-1,18)	0,864
	rs206079	0,802	0,49	0,209	0,50	1,02 (0,86-1,20)	0,834
	rs9534262	0,553	0,48	0,237	0,46	0,92 (0,78-1,09)	0,322
	rs206146	1,000	0,21	0,649	0,24	1,23 (1,01-1,50)	0,039
	rs15869	0,589	0,19	0,476	0,23	1,25 (1,02-1,52)	0,033
<i>FANCD2</i>	rs33917318	0,822	0,11	1,000	0,12	1,11 (0,85-1,45)	0,452
	rs722509	1,000	0,11	0,875	0,12	1,15 (0,94-1,41)	0,182
	rs9811771	1,000	0,12	0,306	0,13	1,11 (0,84-1,45)	0,469
	rs2272125	0,352	0,10	0,883	0,13	1,34 (1,09-1,65)	0,006
<i>FANCE</i>	rs2395626	0,916	0,31	0,539	0,30	0,96 (0,80-1,16)	0,662
	rs760782	0,788	0,48	0,353	0,48	0,99 (0,84-1,17)	0,909
<i>FANCF</i>	rs3740615	0,129	0,15	0,529	0,16	1,09 (0,86-1,38)	0,466
	rs4447177	0,699	0,21	1,000	0,23	1,11 (0,90-1,36)	0,338
<i>FANCG</i>	rs568300	0,409	0,42	0,929	0,39	0,90 (0,76-1,08)	0,257
	rs2258240	0,589	0,21	0,661	0,18	0,83 (0,67-1,03)	0,086
	rs1053318	0,073	0,20	0,527	0,21	1,02 (0,83-1,26)	0,842
<i>FANCI</i>	rs7168941	0,215	0,10	0,404	0,09	0,86 (0,64-1,15)	0,308
	rs8032440	0,538	0,04	0,402	0,03	0,83 (0,52-1,33)	0,438
	rs3087374	0,015	0,09	0,155	0,07	0,71 (0,52-0,98)	0,037
<i>FANCI/BRIP1</i>	rs8080599	0,502	0,16	1,000	0,17	1,12 (0,89-1,41)	0,337
	rs2048718	1,000	0,41	0,727	0,44	1,13 (0,97-1,32)	0,111
	rs4988340	2,26x10⁻²¹	0,21	5,16x10⁻¹⁸	0,24	1,17 (0,98-1,41)	0,090
	rs4988351	0,053	0,23	0,907	0,24	1,05 (0,88-1,26)	0,582
	rs9897928	0,147	0,22	1,000	0,25	1,13 (0,92-1,38)	0,259
	rs1557720	0,014	0,35	1,000	0,36	1,02 (0,85-1,21)	0,857
	rs4986765	0,025	0,28	0,913	0,27	0,96 (0,81-1,14)	0,607
	rs4988357	0,019	0,42	1,000	0,41	0,98 (0,84-1,14)	0,784
	rs7220740	1,000	0,36	0,439	0,34	0,89 (0,74-1,07)	0,196
<i>FANCL</i>	rs3771224	0,610	0,30	0,553	0,32	1,11 (0,92-1,33)	0,266
	rs1011314	0,476	0,30	0,325	0,32	1,11 (0,92-1,33)	0,268
	rs848289	0,685	0,30	0,690	0,31	1,07 (0,89-1,28)	0,471
<i>FANCM</i>	rs8020533	0,661	0,11	0,835	0,12	1,08 (0,83-1,41)	0,577
	rs226984	0,015	0,23	0,557	0,24	1,06 (0,87-1,29)	0,592
	rs226981	0,195	0,45	0,795	0,45	0,98 (0,83-1,16)	0,786
	rs3736772	1,000	0,09	1,000	0,11	1,31 (0,98-1,74)	0,066
<i>FANCN/PALB1</i>	rs8058061	0,200	0,05	0,621	0,05	0,86 (0,59-1,27)	0,456
	rs249941	0,323	0,19	0,683	0,20	1,07 (0,86-1,32)	0,546
	rs16940342	0,547	0,17	0,759	0,17	1,01 (0,81-1,26)	0,939
	rs447529	0,543	0,12	0,181	0,11	0,90 (0,69-1,17)	0,422

Los resultados estadísticamente significativos se indican en negrita rojo. Los resultados marginales se indican en negrita negro.

El análisis estadístico de la distribución de frecuencias alélicas de los SNPs analizados en los genes *FANC* que siguen el HWE en CoCM, desveló asociación a riesgo para cáncer de mama esporádico en el SNP intrónico rs206146 de *FANCD1/BRCA2* (OR=1,23, 95% I.C. 1,01-1,50, $p=0,039$). Asimismo, en el mismo gen encontramos asociación a riesgo para el SNP rs15869, localizado en el extremo 3'UTR (OR=1,25, 95% I.C. 1,02-1,52, $p=0,033$). Ningún otro SNP de los genes pertenecientes a la familia *FANC* presentó asociación a cáncer de mama esporádico, a excepción del SNP rs2272125 cuyo efecto fue descrito en el apartado IV.2.2.

Tras los resultados obtenidos del análisis de las frecuencias alélicas, el estudio de la distribución de los genotipos entre casos y controles (*Tabla 9*) permitió de nuevo la detección de asociación a la enfermedad de los SNPs arriba mencionados para *FANCD1/BRCA2*. Así pues, se pudo constatar una asociación a riesgo estadísticamente significativa para los cambios rs206146 y rs15869 de *FANCD1/BRCA2* (OR=1,23, 95% I.C. 1,01-1,50, $p=0,040$ y OR=1,24, 95% I.C. 1,02-1,52, $p=0,034$, respectivamente). Si bien estos dos SNPs pierden su significación estadística una vez se realiza el análisis multivariante para la corrección por las variables de confusión para cáncer de mama, previamente establecidas, sí que se mantiene el sentido de la asociación.

IV. RESULTADOS

IV.2. Estudio de asociación a cáncer de mama esporádico de genes implicados en la reparación del ADN

Tabla 9: Análisis estadístico de distribución de genotipos en cada grupo de muestras, CME y CoCM y de supervivencia entre CME para la familia de genes *FANCA*.

Gen	SNP	Genotipo Alelo aditivo	Univariante		Multivariate		Supervivencia	
			OR (I.C. 95%)	p-valor	OR (I.C. 95%)	p-valor	OR (I.C. 95%)	p-valor
<i>FANCA</i>	rs2376879	G	0,90 (0,74-1,09)	0,276	0,97 (0,76-1,25)	0,813	0,52 (0,26-1,04)	0,065
	rs7190823	C	0,91 (0,76-1,09)	0,322	0,98 (0,78-1,25)	0,891	0,58 (0,31-1,10)	0,095
	rs1057042	G	0,94 (0,78-1,14)	0,519	0,92 (0,72-1,19)	0,531	0,48 (0,23-0,98)	0,045
	rs2239359	T	1,08 (0,90-1,28)	0,409	0,94 (0,76-1,17)	0,580	0,36 (0,17-0,75)	0,007
	rs3785275	C	0,94 (0,77-1,14)	0,529	0,96 (0,74-1,23)	0,737	0,49 (0,23-1,04)	0,063
	rs7195066	T	0,98 (0,64-1,57)	0,804	1,04 (0,83-1,30)	0,726	0,52 (0,29-0,95)	0,033
	rs2238526	G	1,02 (0,86-1,21)	0,815	0,97 (0,78-1,21)	0,798	0,42 (0,20-0,86)	0,018
	rs1800359	T	0,98 (0,83-1,16)	0,802	1,03 (0,83-1,28)	0,801	2,62 (1,44-4,76)	0,002
<i>FANCB</i>	rs5935816	G	0,98 (0,82-1,17)	0,841	0,95 (0,77-1,17)	0,623	1,16 (0,68-1,99)	0,581
	rs2905223	A	0,98 (0,82-1,17)	0,818	0,97 (0,73-1,30)	0,853	1,17 (0,68-1,99)	0,570
<i>FANCC</i>	rs554879	C	0,99 (0,84-1,18)	0,947	1,15 (0,93-1,42)	0,203	1,17 (0,68-2,02)	0,578
	rs4647554	C	1,01 (0,85-1,21)	0,886	0,96 (0,78-1,19)	0,725	0,77 (0,44-1,33)	0,339
	rs1045276	T	0,94 (0,75-1,17)	0,574	0,90 (0,69-1,18)	0,447	1,95 (1,11-3,43)	0,020
<i>FANCD1/ BRCA2</i>	rs206116	T	1,19 (1,00-1,42)	0,056	1,18 (0,95-1,47)	0,145	0,87 (0,52-1,48)	0,613
	rs3092989	A	0,91 (0,72-1,15)	0,422	0,99 (0,73-1,34)	0,945	1,37 (0,70-2,68)	0,365
	rs1801439	G	1,02 (0,65-1,60)	0,940	0,99 (0,53-1,85)	0,969	0,71 (0,10-5,26)	0,739
	rs1801406	G	0,97 (0,81-1,17)	0,776	1,06 (0,83-1,36)	0,647	0,49 (0,24-0,99)	0,046
	rs206079	G	1,02 (0,87-1,21)	0,798	0,89 (0,72-1,11)	0,306	1,59 (0,93-2,72)	0,092
	rs9534262	C	0,92 (0,78-1,09)	0,323	0,98 (0,79-1,22)	0,884	0,74 (0,44-1,26)	0,271
	rs206146	A	1,23 (1,01-1,50)	0,040	1,31 (0,88-1,45)	0,333	1,00 (0,56-1,78)	0,999
	rs15869	C	1,24 (1,02-1,52)	0,034	1,13 (0,88-1,46)	0,344	1,11 (0,63-1,97)	0,722
<i>FANCD2</i>	rs33917318	C	1,11 (0,85-1,46)	0,451	1,12 (0,80-1,56)	0,506	0,28 (0,07-1,16)	0,080
	rs722509	A	1,15 (0,94-1,42)	0,186	1,06 (0,77-1,46)	0,729	0,52 (0,18-1,46)	0,211
	rs9811771	G	1,11 (0,84-1,46)	0,463	1,09 (0,78-1,53)	0,622	0,57 (0,20-1,65)	0,301
	rs2272125	C	1,36 (1,09-1,68)	0,006	1,27 (0,90-1,79)	0,172	0,40 (0,13-1,30)	0,129
<i>FANCE</i>	rs2395626	G	1,00 (0,85-1,19)	0,963	1,08 (0,88-1,33)	0,469	0,98 (0,56-1,72)	0,954
	rs760782	A	0,96 (0,79-1,15)	0,640	1,01 (0,80-1,27)	0,948	1,04 (0,59-1,85)	0,896
<i>FANCF</i>	rs3740615	T	1,09 (0,87-1,36)	0,476	1,04 (0,79-1,36)	0,782	0,68 (0,32-1,49)	0,340
	rs4447177	G	1,11 (0,90-1,36)	0,335	1,24 (0,97-1,59)	0,091	0,93 (0,50-1,73)	0,815
<i>FANCG</i>	rs568300	T	1,03 (0,83-1,27)	0,823	1,06 (0,81-1,39)	0,674	1,08 (0,59-1,96)	0,813
	rs2258240	T	0,83 (0,67-1,03)	0,084	0,89 (0,68-1,16)	0,395	1,12 (0,61-2,07)	0,720
	rs1053318	T	0,90 (0,76-1,08)	0,250	0,99 (0,80-1,23)	0,918	1,27 (0,73-2,21)	0,405
<i>FANCI</i>	rs7168941	A	0,86 (0,65-1,15)	0,319	0,84 (0,59-1,21)	0,347	1,20 (0,53-2,74)	0,663
	rs8032440	C	0,84 (0,53-1,33)	0,443	0,73 (0,43-1,25)	0,254	0,06 (0,00-22,82)	0,346
<i>FANCI/ BRIP1</i>	rs8080599	T	1,12 (0,89-1,42)	0,331	0,94 (0,70-1,26)	0,681	0,45 (0,16-1,26)	0,128
	rs2048718	T	1,13 (0,97-1,32)	0,121	0,99 (0,80-1,23)	0,960	1,70 (0,93-3,11)	0,088
	rs4988351	G	1,05 (0,87-1,26)	0,635	0,98 (0,77-1,24)	0,836	1,23 (0,69-2,19)	0,490
	rs9897928	A	1,13 (0,92-1,40)	0,243	1,08 (0,84-1,40)	0,528	1,28 (0,70-2,33)	0,419
	rs7220740	A	0,88 (0,74-1,06)	0,186	1,01 (0,80-1,26)	0,953	0,99 (0,51-1,94)	0,987
<i>FANCL</i>	rs3771224	T	1,11 (0,92-1,34)	0,260	1,16 (0,93-1,45)	0,199	1,10 (0,62-1,95)	0,743
	rs1011314	A	1,10 (0,92-1,32)	0,277	1,15 (0,92-1,43)	0,221	0,94 (0,53-1,69)	0,838
	rs848289	A	1,07 (0,89-1,29)	0,467	1,09 (0,87-1,36)	0,459	1,08 (0,61-1,91)	0,785
<i>FANCM</i>	rs8020533	A	1,08 (0,83-1,41)	0,576	1,19 (0,86-1,65)	0,304	0,62 (0,23-1,66)	0,340
	rs226981	A	0,98 (0,83-1,16)	0,788	0,93 (0,76-1,13)	0,447	1,66 (0,96-2,85)	0,068
	rs3736772	C	1,31 (0,98-1,75)	0,066	1,38 (0,98-1,96)	0,069	0,37 (0,09-1,47)	0,159
<i>FANCN/ PALB1</i>	rs8058061	G	0,86 (0,59-1,27)	0,457	1,06 (0,66-1,73)	0,802	0,79 (0,19-3,35)	0,748
	rs249941	T	1,07 (0,86-1,33)	0,543	0,94 (0,72-1,23)	0,667	1,04 (0,54-2,00)	0,919
	rs16940342	G	1,01 (0,81-1,26)	0,940	1,08 (0,82-1,42)	0,592	0,62 (0,29-1,38)	0,239
	rs447529	C	0,90 (0,68-1,17)	0,418	0,97 (0,70-1,33)	0,839	0,85 (0,35-2,08)	0,724

Se ha aplicado un análisis univariante y multivariante para las variables de confusión: edad de diagnóstico o de recogida de muestra, edad de menarquia, número de hijos y estado de menopausia. Los resultados estadísticamente significativos se indican en negrita rojo. Los resultados marginales se indican en negrita negro.

IV.2.3.2. Estudio de asociación de genotipo y fenotipo de los genes *FANC***a. Análisis de supervivencia:**

Se realizó un análisis de supervivencia de las pacientes en función de los genotipos para los SNPs seleccionados en los genes *FANC*. Los resultados indicaron un efecto protector estadísticamente significativo o meramente marginal de los SNPs del gen *FANCA*, algunos de ellos, cambios no sinónimos, a excepción de rs1800359, localizado en el intrón 42, que presenta asociación a un mayor riesgo de fallecimientos. También se detectaron efectos estadísticamente significativos, tanto de protección como de una mayor presencia de fallecimientos, entre las personas portadoras de los alelos menores de los SNPs putativos funcionales rs1801406 (p.Lys1132Lys) en *FANCD1/BRCA2* y rs1045276 (3' corriente abajo) en *FANCC*, respectivamente, y su efecto real en la supervivencia de las pacientes está abierto a discusión (ver *Tabla 9* para más detalle). Estos resultados se encuentran siempre condicionados a la escasa tasa de fallecimientos en la colección de muestras, lo cual limita el poder estadístico de este estudio.

b. Análisis según el tipo de tumor:

Para cada SNP de cada gen seleccionado dentro de la familia *FANC* se realizó además un análisis de genotipo *versus* fenotipo del tumor con el fin de detectar efectos encriptados debido a la heterogeneidad de las muestras de cáncer de mama. El estudio se realizó comparando el efecto del genotipo entre los diferentes grupos de muestras segregadas en función de cada característica tumoral según un modelo de efecto aditivo. Los resultados se describen en la *Tabla 10A* y la *Tabla 10B*. Para aquellas características que presentaron divergencias de frecuencia genotípica entre una categoría y otra, se realizó la estratificación de la muestra acorde a cada característica tumoral y la comparación con la colección de muestras control, con el fin de desvelar la asociación de un SNP con un tipo específico de cáncer de mama.

Los resultados obtenidos indican que los tumores invasivos sólo están asociados al SNP rs4647554, localizado en 3'UTR del gen *FANCC*, presentando un OR de 1,68 estadísticamente significativo (95% I.C. 1,06-2,68, $p=0,028$). Sin embargo, este resultado no mantiene su significación tras realizar el análisis de estratificación de la muestra acorde al tipo de tumor y al genotipo, comparándolo con la totalidad de muestras

control ($p=0,310$). Por otra parte los SNPs rs1045276 a 3' corriente abajo de *FANCC* y rs8058061 a 5' corriente arriba de *FANCN* son más frecuentes en los casos que tienen un tumor *in situ* ($OR=0,48$, 95% I.C. 0,30-0,77, $p=0,002$; y $OR=0,30$, 95% I.C. 0,14-0,64, $p=0,002$, respectivamente). Dichos resultados sí se mantienen en la estratificación y comparación con la colección CoCM, con p-valores de $p=0,009$ y $p=0,013$, respectivamente.

Acorde al grado histológico del tumor, se establecieron dos asociaciones estadísticamente significativas entre tumores de bajo grado y las variantes rs2395626 en *FANCE* y rs8982440 en *FANCI*, ambos intrónicos. Los valores de OR estimados fueron 0,66 (95% I.C. 0,47-0,93, $p=0,019$) y 0,41 (95% I.C. 0,17-0,98, $p=0,045$), respectivamente, aunque en ambos casos dicho resultado no fue consistente en el análisis de estratificación ($p=0,079$ y $p=0,331$, respectivamente).

Por otra parte, tumores de mayor tamaño que 2 cm en su diámetro mayor se han visto relacionados con tres SNPs en *FANCD2*, rs33917318 (5' corriente arriba del gen), rs722509 y rs9811771 (ambos intrónicos). Éstos presentan unos valores de riesgo estimado según: 1,71 (95% I.C. 1,09-2,69, $p=0,019$), 1,63 (95% I.C. 1,07-2,50, $p=0,024$) y 1,63 (95% I.C. 1,02-2,61, $p=0,040$), respectivamente. Sin embargo, estas asociaciones con la enfermedad por parte del SNP y acorde al tamaño del tumor sólo se mantuvieron en el estudio de estratificación de la muestra en el caso de rs722509, con un p-valor de 0,039. Asimismo, otra variante, rs568300, localizada a 5' corriente arriba del gen *FANCG*, se ha visto relacionada con tumores >2 cm, establecido su OR estimado en 1,42 (95% I.C. 1,01-2,01), con una significación marginal de $p=0,044$, no conservada tras en estudio de estratificación.

c. Análisis de los marcadores tumorales:

El estudio de los marcadores tumorales no arrojó luz en la relación entre tumores ER positivos y los diferentes SNPs testados, mientras que tumores positivos para PR presentaron asociaciones estadísticamente significativas con los SNPs rs3785275 en *FANCA* (intrónico, $OR=1,65$, 95% I.C. 1,12-2,44, $p=0,012$), rs7220740 en *FANCI/BRIP1* (3' corriente abajo, $OR=1,60$, 95% I.C. 1,14-2,27, $p=0,007$) y rs447529 en *FANCN/PALB2* (intrónico, $OR=1,86$, 95% I.C. 1,06-3,25, $p=0,030$). Los tres SNPs analizados sobre una

muestra estratificada según el estatus de PR se asociaron a la enfermedad con p-valores significativos: $p=0,012$, $p=0,039$ y $p=0,039$, respectivamente.

d. Análisis del estado de metástasis:

Los análisis de la relación entre los SNPs testados y el riesgo de presentar metástasis a ganglios linfáticos cercanos o metástasis a tejidos distantes del tejido mamario, permitieron estimar ciertas asociaciones estadísticamente significativas, si bien con sentidos contrapuestos. Así, el SNP rs2395626 localizado en *FANCE*, ya anteriormente citado para otra asociación clínica en este apartado, se encuentra con mayor frecuencia entre pacientes sin metástasis a ganglios linfáticos adyacentes a la mama, según $OR=0,74$ (95% I.C. 0,77-1,00), aunque con un carácter marginal ($p=0,048$) no observado en el estudio de estratificación ($p=0,286$). No ocurre lo mismo con el SNP intrónico rs8032440 en *FANCI*, que presenta asociación con pacientes con ganglios linfáticos no metastatizados ($OR=0,37$; 95% I.C. 0,14-0,98; $p=0,045$) y asociación con la enfermedad entre estos pacientes ($p=0,048$), aunque con un carácter marginal. Sin embargo, otro SNP localizado en *FANCI*, rs7168941, también de localización intrónica dentro del gen, está asociado a metástasis de ganglios linfáticos: $OR=1,74$ (95% I.C. 1,06-2,87; $p=0,030$), aunque dicho resultado significativo se pierde tras la estratificación de la muestra ($p=0,089$). En los casos en que se analizó la metástasis a distancia, sólo se pudo detectar asociación en los SNPs de *FANCA*: rs1800359, intrónico ($OR=1,64$, 95% I.C. 1,28-5,41, $p=0,008$), y rs2239359, cambio no sinónimo p.Ser501Gly ($OR=0,11$, 95% I.C. 0,03-0,48, $p=0,003$). Ambas asociaciones de riesgo y protección frente a metástasis fueron confirmadas en su implicación en el desarrollo de la enfermedad tras el estudio de estratificación de la muestra, con p-valores de $p=0,009$ y $p=0,005$, respectivamente.

IV. RESULTADOS

IV.2. Estudio de asociación a cáncer de mama esporádico de genes implicados en la reparación del ADN

Tabla 10A: Análisis estadístico de los genotipos de los genes de la familia *FANC* frente al tipo de tumor.

Gen	SNP	Tumor invasivo		Grado de tumor > 1		Tamaño de tumor > 2cm	
		OR (I.C. 95%)	p-valor	OR (I.C. 95%)	p-valor	OR (I.C. 95%)	p-valor
<i>FANCA</i>	rs2376879	1,11 (0,69-1,80)	0,666	1,10 (0,75-1,60)	0,632	1,28 (0,93-1,76)	0,124
	rs7190823	0,88 (0,56-1,39)	0,591	1,10 (0,77-1,58)	0,601	1,03 (0,76-1,40)	0,840
	rs1057042	0,91 (0,56-1,48)	0,705	1,09 (0,74-1,60)	0,661	1,10 (0,79-1,53)	0,556
	rs2239359	0,86 (0,57-1,31)	0,484	0,89 (0,64-1,24)	0,482	0,98 (0,74-1,31)	0,907
	rs3785275	0,93 (0,56-1,51)	0,768	1,13 (0,76-1,67)	0,539	1,15 (0,83-1,61)	0,402
	rs7195066	0,97 (0,64-1,46)	0,878	0,99 (0,71-1,38)	0,937	0,90 (0,68-1,20)	0,488
	rs2238526	1,07 (0,68-1,70)	0,765	0,99 (0,70-1,41)	0,966	1,13 (0,84-1,53)	0,430
<i>FANCB</i>	rs1800359	1,02 (0,68-1,53)	0,940	1,08 (0,77-1,50)	0,671	1,00 (0,76-1,33)	0,976
	rs5935816	1,26 (0,81-1,97)	0,314	0,90 (0,64-1,27)	0,560	1,05 (0,79-1,41)	0,729
	rs2905223	1,26 (0,81-1,98)	0,308	0,92 (0,65-1,30)	0,635	1,06 (0,79-1,42)	0,703
<i>FANCC</i>	rs554879	0,87 (0,56-1,33)	0,507	0,76 (0,54-1,06)	0,108	0,84 (0,63-1,12)	0,227
	rs4647554	1,68 (1,06-2,68)	0,028	1,17 (0,83-1,66)	0,372	1,23 (0,92-1,64)	0,159
	rs1045276	0,48 (0,30-0,77)	0,002	1,58 (0,98-2,54)	0,060	0,88 (0,61-1,27)	0,502
<i>FANCD1/BRCA2</i>	rs206116	0,99 (0,65-1,51)	0,957	0,77 (0,56-1,07)	0,117	0,80 (0,60-1,08)	0,141
	rs3092989	1,08 (0,60-1,98)	0,791	1,03 (0,64-1,66)	0,893	1,23 (0,83-1,84)	0,306
	rs1801439	1,08 (0,31-3,72)	0,908	1,24 (0,44-3,48)	0,682	1,42 (0,65-3,12)	0,381
	rs1801406	1,04 (0,65-1,68)	0,864	0,75 (0,52-1,09)	0,130	0,76 (0,54-1,05)	0,095
	rs206079	1,04 (0,68-1,58)	0,871	1,14 (0,82-1,59)	0,433	1,03 (0,78-1,37)	0,833
	rs9534262	0,93 (0,61-1,42)	0,737	1,08 (0,78-1,52)	0,638	1,24 (0,93-1,66)	0,144
	rs206146	1,46 (0,86-2,46)	0,158	0,72 (0,50-1,04)	0,078	0,88 (0,64-1,22)	0,445
<i>FANCD2</i>	rs15869	1,34 (0,80-2,26)	0,271	0,70 (0,49-1,01)	0,059	0,97 (0,70-1,35)	0,874
	rs33917318	0,91 (0,47-1,78)	0,787	1,11 (0,63-1,93)	0,722	1,71 (1,09-2,69)	0,019
	rs722509	0,97 (0,50-1,86)	0,916	1,16 (0,68-1,99)	0,580	1,63 (1,07-2,50)	0,024
	rs9811771	0,64 (0,34-1,20)	0,166	1,17 (0,64-2,12)	0,610	1,63 (1,02-2,61)	0,040
<i>FANCE</i>	rs2272125	0,84 (0,46-1,52)	0,561	1,18 (0,71-1,98)	0,519	1,45 (0,96-2,21)	0,080
	rs2395626	0,73 (0,47-1,13)	0,158	0,66 (0,47-0,93)	0,019	0,90 (0,67-1,21)	0,472
	rs760782	0,91 (0,57-1,44)	0,674	0,73 (0,51-1,04)	0,083	0,88 (0,64-1,22)	0,447
<i>FANCF</i>	rs3740615	0,76 (0,45-1,26)	0,279	0,96 (0,62-1,49)	0,863	0,78 (0,53-1,14)	0,197
	rs4447177	0,93 (0,57-1,52)	0,770	1,06 (0,71-1,58)	0,781	1,01 (0,72-1,42)	0,945
<i>FANCG</i>	rs568300	1,00 (0,57-1,73)	0,994	1,26 (0,82-1,94)	0,300	1,42 (1,01-2,01)	0,044
	rs2258240	1,15 (0,67-1,99)	0,610	0,89 (0,59-1,36)	0,594	0,83 (0,57-1,20)	0,320
	rs1053318	1,19 (0,76-1,84)	0,452	1,10 (0,78-1,55)	0,596	1,18 (0,88-1,57)	0,270
<i>FANCI</i>	rs7168941	1,31 (0,58-2,94)	0,518	1,05 (0,56-1,99)	0,870	1,14 (0,68-1,92)	0,618
	rs8032440	1,08 (0,31-3,74)	0,900	0,41 (0,17-0,98)	0,045	0,53 (0,22-1,29)	0,159
<i>FANCI/BRIP1</i>	rs8080599	1,07 (0,61-1,88)	0,806	1,02 (0,65-1,61)	0,921	0,82 (0,56-1,21)	0,321
	rs2048718	1,15 (0,76-1,75)	0,499	0,97 (0,70-1,34)	0,848	0,93 (0,70-1,23)	0,601
	rs4988351	0,98 (0,61-1,58)	0,936	1,10 (0,75-1,61)	0,630	1,14 (0,83-1,57)	0,411
	rs9897928	0,97 (0,60-1,59)	0,912	1,11 (0,75-1,63)	0,609	1,12 (0,81-1,55)	0,494
	rs7220740	1,07 (0,98-1,67)	0,784	0,74 (0,52-1,05)	0,096	0,84 (0,62-1,15)	0,278
<i>FANCL</i>	rs3771224	1,01 (0,64-1,60)	0,968	0,91 (0,62-1,31)	0,614	0,81 (0,59-1,11)	0,196
	rs1011314	0,91 (0,59-1,40)	0,666	0,92 (0,65-1,30)	0,633	0,95 (0,71-1,28)	0,732
	rs848289	1,05 (0,67-1,67)	0,826	0,98 (0,68-1,42)	0,918	0,82 (0,60-1,12)	0,204
<i>FANCM</i>	rs8020533	0,96 (0,51-1,79)	0,888	0,80 (0,49-1,29)	0,354	1,20 (0,79-1,82)	0,393
	rs226981	1,14 (0,75-1,74)	0,530	0,98 (0,70-1,36)	0,884	0,87 (0,65-1,15)	0,328
	rs3736772	1,26 (0,62-2,58)	0,526	0,88 (0,53-1,46)	0,628	1,29 (0,83-2,01)	0,263
<i>FANCN/PALB1</i>	rs8058061	0,30 (0,14-0,64)	0,002	1,03 (0,44-2,39)	0,946	0,55 (0,26-1,18)	0,124
	rs249941	1,39 (0,79-2,43)	0,251	1,20 (0,78-1,83)	0,411	1,13 (0,80-1,61)	0,482
	rs16940342	0,93 (0,55-1,58)	0,783	0,69 (0,46-1,04)	0,075	1,18 (0,82-1,70)	0,388
	rs447529	1,10 (0,56-2,19)	0,778	1,01 (0,59-1,72)	0,971	0,81 (0,50-1,32)	0,394

Análisis estadístico del efecto aditivo del genotipo (en su alelo menor) para la familia de genes *FANC* entre los diferentes grupos de muestras segregadas en función de las características tumorales: tumor invasivo *versus* tumor *in situ*, tumor de grado >1 *versus* tumor de grado ≤1, tumor de tamaño >2 cm *versus* tumor de tamaño ≤2 cm. Los resultados estadísticamente significativos se indican en negrita rojo. Los resultados marginales se indican en negrita negro.

Tabla 10B: Análisis estadístico de los genotipos de los genes de la familia FANC frente al estado de los marcadores tumorales y de metástasis.

Gen	SNP	ER Positivo		PR Positivo		Metástasis a ganglio		Otras metástasis	
		OR (I.C. 95%)	p-valor	OR (I.C. 95%)	p-valor	OR (I.C. 95%)	p-valor	OR (I.C. 95%)	p-valor
FANCA	rs2376879	0,97 (0,64-1,48)	0,891	1,15 (0,80-1,64)	0,456	1,17 (0,85-1,61)	0,336	0,92 (0,43-1,98)	0,835
	rs7190823	0,84 (0,56-1,25)	0,390	1,18 (0,84-1,66)	0,347	1,05 (0,77-1,43)	0,751	0,71 (0,33-1,53)	0,385
	rs1057042	0,97 (0,63-1,49)	0,890	1,44 (0,99-2,10)	0,059	1,07 (0,76-1,49)	0,704	0,96 (0,44-2,10)	0,915
	rs2239359	1,06 (0,73-1,55)	0,764	0,88 (0,64-1,19)	0,400	0,89 (0,66-1,19)	0,434	0,11 (0,03-0,48)	0,003
	rs3785275	1,07 (0,68-1,66)	0,778	1,65 (1,12-2,44)	0,012	1,07 (0,76-1,51)	0,683	1,07 (0,49-2,33)	0,860
	rs7195066	0,95 (0,66-1,38)	0,804	1,05 (0,77-1,43)	0,769	1,01 (0,76-1,35)	0,957	0,54 (0,26-1,10)	0,087
	rs2238526	0,91 (0,61-1,36)	0,638	1,17 (0,83-1,64)	0,366	1,01 (0,75-1,38)	0,933	0,75 (0,35-1,62)	0,468
	rs1800359	1,01 (0,70-1,46)	0,963	0,96 (0,71-1,30)	0,785	1,07 (0,80-1,42)	0,660	2,64 (1,28-5,41)	0,008
FANCB	rs5935816	0,90 (0,61-1,31)	0,566	0,85 (0,62-1,18)	0,334	1,03 (0,77-1,37)	0,864	0,97 (0,49-1,94)	0,939
	rs2905223	0,91 (0,63-1,33)	0,626	0,87 (0,63-1,20)	0,396	1,05 (0,79-1,41)	0,731	0,98 (0,49-1,94)	0,947
FANCC	rs554879	0,85 (0,58-1,23)	0,386	0,88 (0,64-1,21)	0,422	0,78 (0,59-1,05)	0,096	0,77 (0,39-1,53)	0,463
	rs4647554	1,02 (0,69-1,49)	0,933	1,24 (0,89-1,74)	0,203	1,24 (0,92-1,65)	0,156	1,40 (0,72-2,72)	0,323
	rs1045276	1,22 (0,75-1,97)	0,429	0,79 (0,53-1,18)	0,252	1,02 (0,71-1,48)	0,909	0,83 (0,35-1,99)	0,678
FANCD1/ BRCA2	rs206116	1,16 (0,80-1,70)	0,434	1,07 (0,78-1,48)	0,666	0,99 (0,74-1,32)	0,933	1,07 (0,55-2,06)	0,848
	rs3092989	0,89 (0,54-1,48)	0,655	1,29 (0,82-2,01)	0,271	0,81 (0,54-1,24)	0,333	1,28 (0,54-3,07)	0,578
	rs1801439	1,27 (0,41-3,94)	0,682	2,01 (0,76-5,32)	0,159	1,86 (0,83-4,19)	0,134	0,84 (0,11-6,54)	0,869
	rs1801406	1,08 (0,71-1,64)	0,722	0,90 (0,63-1,29)	0,571	1,09 (0,79-1,51)	0,593	1,40 (0,69-2,84)	0,357
	rs206079	0,91 (0,62-1,33)	0,631	0,97 (0,70-1,34)	0,847	1,08 (0,81-1,44)	0,600	0,88 (0,45-1,70)	0,694
	rs9534262	1,04 (0,71-1,51)	0,841	1,04 (0,75-1,43)	0,823	0,86 (0,64-1,15)	0,294	1,58 (0,78-3,21)	0,206
	rs206146	1,02 (0,67-1,57)	0,915	0,95 (0,66-1,37)	0,794	1,02 (0,74-1,41)	0,895	0,56 (0,23-1,36)	0,199
	rs15869	0,92 (0,60-1,41)	0,695	0,93 (0,65-1,34)	0,693	1,07 (0,77-1,48)	0,692	0,42 (0,15-1,19)	0,101
FANCD2	rs33917318	0,89 (0,50-1,60)	0,703	1,11 (0,67-1,86)	0,686	0,80 (0,49-1,31)	0,383	0,45 (0,11-1,94)	0,286
	rs722509	1,10 (0,60-2,00)	0,759	1,15 (0,71-1,89)	0,567	0,81 (0,51-1,28)	0,370	0,40 (0,10-1,71)	0,218
	rs9811771	0,96 (0,51-1,78)	0,888	1,35 (0,79-2,30)	0,277	0,85 (0,52-1,41)	0,534	0,55 (0,12-2,46)	0,435
	rs2272125	0,90 (0,52-1,56)	0,708	1,03 (0,65-1,65)	0,893	0,74 (0,47-1,17)	0,202	0,38 (0,09-1,61)	0,189
FANCE	rs2395626	0,98 (0,68-1,42)	0,927	1,06 (0,77-1,45)	0,727	0,74 (0,55-1,00)	0,048	0,64 (0,31-1,30)	0,213
	rs760782	1,02 (0,68-1,52)	0,943	1,07 (0,75-1,50)	0,721	0,83 (0,61-1,14)	0,253	0,80 (0,37-1,72)	0,564
FANCF	rs3740615	1,02 (0,64-1,61)	0,941	0,94 (0,64-1,38)	0,745	0,88 (0,61-1,29)	0,518	0,75 (0,29-1,94)	0,558
	rs4447177	1,16 (0,74-1,83)	0,513	1,36 (0,92-2,00)	0,122	0,98 (0,70-1,37)	0,907	1,62 (0,80-3,28)	0,179
FANCG	rs568300	1,28 (0,78-2,12)	0,332	1,15 (0,76-1,73)	0,504	1,02 (0,72-1,44)	0,930	1,93 (0,94-4,00)	0,075
	rs2258240	0,72 (0,46-1,14)	0,159	0,94 (0,64-1,40)	0,766	1,11 (0,77-1,59)	0,579	0,89 (0,36-2,19)	0,804
	rs1053318	0,94 (0,64-1,37)	0,732	1,06 (0,77-1,46)	0,745	1,11 (0,83-1,49)	0,472	1,37 (0,70-2,66)	0,355
FANCI	rs7168941	1,81 (0,83-3,96)	0,137	1,24 (0,68-2,26)	0,481	1,74 (1,06-2,87)	0,030	1,55 (0,53-4,52)	0,424
	rs8032440	2,32 (0,52-10,46)	0,272	1,52 (0,56-4,11)	0,407	0,37 (0,14-0,98)	0,045	-	-
	rs8080599	1,28 (0,76-2,14)	0,354	1,09 (0,72-1,63)	0,687	0,72 (0,48-1,07)	0,101	0,76 (0,29-2,00)	0,573
FANCI/ BRIP1	rs2048718	0,84 (0,58-1,20)	0,334	0,89 (0,66-1,20)	0,442	0,903 (0,68-1,20)	0,481	1,33 (0,69-2,56)	0,402
	rs4988351	1,09 (0,71-1,67)	0,702	0,99 (0,70-1,41)	0,970	0,97 (0,70-1,35)	0,855	1,18 (0,57-2,44)	0,658
	rs9897928	1,07 (0,70-1,65)	0,754	0,89 (0,62-1,27)	0,509	1,05 (0,75-1,46)	0,776	1,00 (0,45-2,23)	0,996
	rs7220740	1,44 (0,97-2,16)	0,074	1,60 (1,14-2,27)	0,007	1,17 (0,86-1,59)	0,317	0,81 (0,38-1,70)	0,574
FANCL	rs3771224	1,09 (0,72-1,65)	0,680	0,95 (0,67-1,34)	0,764	0,97 (0,71-1,32)	0,838	1,02 (0,48-2,15)	0,962
	rs1011314	1,24 (0,83-1,85)	0,285	1,23 (0,89-1,69)	0,209	0,95 (0,71-1,28)	0,744	0,77 (0,37-1,59)	0,473
	rs848289	0,99 (0,66-1,49)	0,968	0,94 (0,67-1,32)	0,707	0,98 (0,72-1,34)	0,919	0,90 (0,44-1,86)	0,783
FANCM	rs8020533	1,02 (0,59-1,79)	0,935	0,93 (0,58-1,50)	0,770	1,24 (0,82-1,88)	0,308	1,72 (0,75-3,94)	0,199
	rs226981	1,15 (0,80-1,66)	0,441	1,27 (0,93-1,72)	0,128	1,09 (0,82-1,45)	0,541	1,96 (1,00-3,84)	0,051
	rs3736772	1,04 (0,57-1,89)	0,912	1,01 (0,60-1,71)	0,962	1,39 (0,89-2,15)	0,146	1,99 (0,85-4,66)	0,114
FANCN/ PALB1	rs8058061	1,36 (0,57-3,23)	0,493	2,16 (0,98-4,77)	0,056	0,87 (0,41-1,84)	0,718	0,51 (0,07-3,88)	0,511
	rs249941	1,14 (0,71-1,84)	0,591	0,96 (0,66-1,41)	0,844	1,18 (0,83-1,66)	0,363	1,12 (0,50-2,46)	0,804
	rs16940342	1,26 (0,76-2,10)	0,371	1,15 (0,77-1,73)	0,498	0,76 (0,52-1,12)	0,169	1,61 (0,75-3,46)	0,226
	rs447529	1,60 (0,82-3,13)	0,165	1,86 (1,06-3,25)	0,030	0,68 (0,42-1,12)	0,129	0,94 (0,31-2,82)	0,909

Análisis estadístico del efecto aditivo del genotipo (en su alelo menor) para la familia de genes FANC entre los diferentes grupos de muestras segregadas en función de las características tumorales: tumor ER positivo *versus* tumor ER negativo, tumor PR positivo *versus* tumor PR negativo, tumor metastatizado a ganglio linfático axilar *versus* tumor no metastatizado a ganglio linfático axilar, y tumor metastatizado a otros tejidos distales *versus* tumor no metastatizado a otros tejidos distales. Los resultados estadísticamente significativos se indican en negrita rojo. Los resultados marginales se indican en negrita negro.

IV.2.4. Estudio de otros genes relacionados con la familia de genes *FANC*

A raíz del estudio de todos los genes pertenecientes a la familia *FANC*, se consideró de interés estudiar de igual modo alguno de los genes cuyas proteínas intervienen de algún modo en la función de las proteínas *FANC*. Así, se estudió un total de 6 genes: *ATM*, *ATR*, *BRCA1*, *H2AX*, *OGG1* y *USP1*.

IV.2.4.1. Estudio de asociación de genotipado de los genes relacionados con los genes *FANC*

El análisis de la distribución de los tres tipos de genotipos (homocigoto normal, heterocigoto y homocigoto raro) de los genes relacionados con los genes *FANC* entre CME y CoCM permitió identificar un total de 5 SNPs que se desviaban de manera estadísticamente significativa de HWE ($p < 0,05$) en CoCM y que se descartaron del estudio estadístico posterior. De éstos, tres presentaron esta desviación también en CME, y otros tres se desviaban de HWE exclusivamente entre CME y se siguieron considerando en estudios posteriores, al igual que aquellos controles con una estadística marginal de desviación de HWE (*Tabla 11*).

El análisis estadístico de la distribución de frecuencias alélicas de los SNPs analizados de los genes relacionados con los genes *FANC* que siguen el HWE en controles, desveló asociación a riesgo para el cáncer de mama en el SNP intrónico rs637064 de *ATM* ($OR=1,19$, 95% I.C. 1,01-1,41, $p=0,037$). Sin embargo un SNP intrónico en el gen *BRCA1*, rs799923, presentó un efecto protector para el cáncer de mama estadísticamente significativo ($OR=0,82$, 95% I.C. 0,67-1,00, $p=0,048$). Ningún otro SNP de los genes relacionados con los genes *FANC* estudiados presentó asociación a cáncer de mama esporádico (*Tabla 11*).

Tabla 11: Análisis estadístico de HWE, indicación de MAF en cada grupo de muestras y análisis estadístico de la distribución de frecuencias alélicas entre CoCM y CME para los genes relacionados con la familia de genes *FANC*.

Gen	SNP	CoCM		CME		OR (I.C. 95%)	p-valor
		p-valor HWE	MAF	p-valor HWE	MAF		
<i>ATM</i>	rs228589	0,521	0,38	0,088	0,40	1,08 (0,91-1,29)	0,366
	rs189037	0,084	0,38	0,039	0,40	1,08 (0,91-1,28)	0,380
	rs625120	0,148	0,40	0,294	0,40	1,04 (0,89-1,21)	0,661
	rs228590	0,783	0,38	0,178	0,40	1,08 (0,91-1,29)	0,379
	rs228599	0,709	0,37	0,271	0,38	1,06 (0,89-1,26)	0,509
	rs599164	0,552	0,36	0,319	0,38	1,10 (0,93-1,30)	0,284
	rs637064	0,736	0,37	0,293	0,42	1,19 (1,01-1,41)	0,037
	rs1800058	0,042	0,01	1,000	0,01	1,59 (0,72-3,53)	0,246
	rs647681	0,865	0,37	0,371	0,39	1,09 (0,92-1,29)	0,306
	rs1801516	0,738	0,14	0,011	0,13	0,94 (0,74-1,19)	0,593
	rs660429	0,713	0,38	0,127	0,40	1,08 (0,91-1,29)	0,373
	rs227060	0,232	0,38	0,165	0,37	0,96 (0,81-1,14)	0,639
	rs227064	0,780	0,37	0,314	0,38	1,05 (0,88-1,25)	0,592
	rs172896	4,77x10⁻²⁰	0,37	5,93x10⁻¹⁹	0,37	-	-
	rs664143	0,641	0,37	0,274	0,38	1,07 (0,90-1,28)	0,447
	rs170548	0,272	0,38	0,094	0,37	0,94 (0,79-1,12)	0,475
	rs4585	0,852	0,38	0,027	0,39	1,05 (0,88-1,26)	0,564
<i>ATR</i>	rs6782400	0,854	0,41	0,538	0,40	0,98 (0,82-1,16)	0,791
	rs7636909	0,147	0,44	0,171	0,46	1,08 (0,91-1,29)	0,356
	rs12635931	0,078	0,13	0,264	0,13	1,00 (0,78-1,29)	0,991
<i>BRCA1</i>	rs799906	0,018	0,36	1,000	0,38	-	-
	rs799923	0,720	0,22	0,137	0,19	0,82 (0,67-1,00)	0,048
	rs1799949	2,89x10⁻⁹	0,29	3,60x10⁻⁵	0,32	-	-
	rs16941	0,122	0,34	0,461	0,37	1,17 (0,98-1,40)	0,084
	rs8176265	0,515	0,34	0,787	0,36	1,11 (0,94-1,32)	0,221
	rs8176318	0,458	0,34	0,857	0,37	1,13 (0,95-1,34)	0,164
<i>H2AX</i>	rs8551	0,932	0,46	0,346	0,47	1,05 (0,91-1,24)	0,548
	rs7759	0,532	0,39	0,420	0,39	1,02 (0,86-1,21)	0,791
	rs7350	0,654	0,38	0,173	0,38	1,01 (0,85-1,19)	0,946
<i>OGG1</i>	rs159153	0,373	0,31	0,688	0,33	1,10 (0,91-1,32)	0,326
	rs125701	0,869	0,15	0,207	0,16	1,08 (0,86-1,36)	0,520
	rs3219008	0,190	0,23	0,061	0,22	0,94 (0,77-1,15)	0,554
	rs1052133	0,259	0,22	0,157	0,21	0,97 (0,79-1,19)	0,750
<i>USP1</i>	rs9436223	0,466	0,29	0,342	0,29	0,97 (0,81-1,17)	0,771
	rs614665	0,000	0,00	0,000	0,00	0,34 (0,01-8,35)	1,372
	rs583609	0,601	0,29	0,338	0,28	0,99 (0,82-1,19)	0,876

Los resultados estadísticamente significativos se indican en negrita rojo. Los resultados marginales se indican en negrita negro.

Nuevamente, los resultados de la distribución de genotipos en casos y controles coincidieron con los resultados de las frecuencias alélicas, estableciendo una asociación a protección frente a la enfermedad para el SNP rs637064 de *ATM* (OR=1,19, 95% I.C. 1,01-1,40; p=0,041), efecto que tras una corrección por análisis multivariante para las variables de confusión descritas previamente, se mantiene pero sin significación estadística. Por otra parte, de nuevo, el resultado para *BRCA1* descrito en el apartado anterior se ve reforzado por una distribución de genotípica variable en rs799923 (OR=0,81, 95% I.C. 0,66-0,99, p=0,041), que se convierte en una significación marginal

IV. RESULTADOS

IV.2. Estudio de asociación a cáncer de mama esporádico de genes implicados en la reparación del ADN

($p=0,056$) aunque con el mismo efecto que en el análisis univariante (ver *Tabla 12* para más detalle).

Tabla 12: Análisis estadístico de distribución de genotipos en cada grupo de muestras, CME y CoCM y de supervivencia entre CME para los genes relacionados con la familia de genes *FANC*.

Gen	SNP	Genotipo Alelo aditivo	Univariante		Multivariante		Supervivencia	
			OR (I.C. 95%)	p-valor	OR (I.C. 95%)	p-valor	OR (95%CI)	p-valor
ATM	rs228589	T	1,08 (0,91-1,28)	0,378	1,08 (0,88-1,33)	0,447	0,80 (0,47-1,36)	0,402
	rs189037	G	1,07 (0,91-1,27)	0,399	1,14 (0,93-1,39)	0,211	0,82 (0,49-1,37)	0,446
	rs625120	T	1,01 (0,87-1,18)	0,908	1,06 (0,87-1,29)	0,572	1,08 (0,63-1,84)	0,783
	rs228590	T	1,08 (0,91-1,28)	0,406	1,11 (0,90-1,36)	0,336	1,03 (0,61-1,72)	0,922
	rs228599	A	1,06 (0,89-1,25)	0,538	1,12 (0,91-1,38)	0,288	1,07 (0,64-1,79)	0,793
	rs599164	A	1,09 (0,92-1,29)	0,324	1,12 (0,91-1,37)	0,293	1,04 (0,62-1,75)	0,873
	rs637064	C	1,19 (1,01-1,40)	0,041	1,11 (0,91-1,37)	0,302	1,07 (0,63-1,81)	0,797
	rs647681	A	1,09 (0,92-1,28)	0,321	1,11 (0,91-1,36)	0,300	1,04 (0,63-1,73)	0,895
	rs1801516	A	0,95 (0,75-1,20)	0,677	0,92 (0,69-1,22)	0,546	1,19 (0,56-2,52)	0,655
	rs660429	C	1,08 (0,91-1,28)	0,401	1,09 (0,89-1,34)	0,409	0,97 (0,58-1,62)	0,893
	rs227060	T	0,96 (0,81-1,14)	0,625	0,95 (0,77-1,17)	0,617	1,02 (0,60-1,74)	0,931
	rs227064	G	1,05 (0,88-1,24)	0,620	1,10 (0,89-1,35)	0,369	1,00 (0,60-1,68)	0,990
	rs664143	C	1,07 (0,90-1,27)	0,475	1,13 (0,92-1,39)	0,249	1,07 (0,64-1,78)	0,804
	rs170548	C	0,94 (0,79-1,11)	0,467	0,94 (0,76-1,16)	0,537	0,92 (0,53-1,61)	0,779
ATR	rs4585	T	1,05 (0,88-1,25)	0,598	1,09 (0,88-1,34)	0,442	1,08 (0,64-1,84)	0,768
	rs6782400	C	0,98 (0,82-1,16)	0,783	0,93 (0,75-1,16)	0,537	0,59 (0,32-1,07)	0,080
	rs7636909	G	1,09 (0,92-1,31)	0,321	1,17 (0,93-1,45)	0,175	1,37 (0,79-2,38)	0,263
	rs12635931	C	1,00 (0,78-1,29)	0,978	0,96 (0,70-1,32)	0,818	0,40 (0,15-1,23)	0,083
BRCA1	rs799923	A	0,81 (0,66-0,99)	0,041	0,77 (0,59-1,01)	0,056	0,72 (0,34-1,54)	0,397
	rs16941	G	1,17 (0,98-1,39)	0,082	1,27 (1,01-1,58)	0,037	1,34 (0,75-2,36)	0,322
	rs8176265	A	1,12 (0,94-1,32)	0,211	1,19 (0,96-1,48)	0,111	1,45 (0,86-2,46)	0,162
	rs8176318	T	1,13 (0,96-1,34)	0,153	1,23 (0,99-1,52)	0,065	1,48 (0,87-2,51)	0,146
H2AX	rs8551	C	1,05 (0,89-1,24)	0,575	1,06 (0,87-1,30)	0,571	0,73 (0,43-1,23)	0,232
	rs7759	A	1,02 (0,86-1,21)	0,829	1,05 (0,85-1,30)	0,635	0,92 (0,54-1,58)	0,764
	rs7350	C	1,00 (0,84-1,18)	0,977	1,04 (0,85-1,29)	0,696	0,95 (0,55-1,63)	0,844
OGG1	rs159153	C	1,10 (0,91-1,32)	0,324	1,18 (0,95-1,48)	0,141	1,41 (0,83-2,38)	0,203
	rs125701	A	1,08 (0,86-1,37)	0,514	1,26 (0,94-1,68)	0,120	1,50 (0,72-3,11)	0,278
	rs3219008	G	0,94 (0,77-1,15)	0,556	0,89 (0,70-1,14)	0,365	0,92 (0,51-1,66)	0,778
	rs1052133	G	0,97 (0,79-1,19)	0,750	0,93 (0,72-1,19)	0,564	0,89 (0,48-1,67)	0,722
USP1	rs9436223	T	0,97 (0,81-1,17)	0,771	1,07 (0,85-1,34)	0,579	1,28 (0,71-2,33)	0,411
	rs583609	C	0,99 (0,82-1,19)	0,876	1,07 (0,85-1,34)	0,565	1,21 (0,67-2,19)	0,519

Se ha aplicado un análisis univariante y multivariante para las variables de confusión: edad de diagnóstico o de recogida de muestra, edad de menarquia, número de hijos, estado de menopausia. Los resultados estadísticamente significativos se indican en negrita rojo. Los resultados marginales se indican en negrita negro.

IV.2.4.2. Estudio de asociación de genotipo y fenotipo de los genes relacionados con los genes *FANC*

a. Análisis de supervivencia:

El análisis de supervivencia de las pacientes en función de los genotipos para los SNPs seleccionados en los genes relacionados con los genes *FANC*, expuesto en la *Tabla 12*, no evidenció ningún efecto estadísticamente significativo para ninguno de los SNPs analizados.

b. Análisis según el tipo de tumor:

Al igual que ocurrió con el estudio de la familia de genes *FANC*, se realizó un análisis de genotipo *versus* fenotipo del tumor para el grupo de genes relacionados con la familia *FANC*, siguiendo las directrices del análisis estadístico expuesto en el apartado III.5.4. Los resultados de dicho análisis se presentan en la *Tabla 13A* y la *Tabla 13B*.

Los resultados obtenidos indican que los tumores *in situ* están asociados a tres variantes en *BRCA1*: rs16941, un cambio no sinónimo (p.Gly1038Glu) del exón 10 del gen (OR=0,62, 95% I.C. 0,40-0,96, p=0,030), rs8176265, intrónico (OR=0,65, 95% I.C. 0,43-0,99, p=0,046), y rs8176318, localizado en 3'UTR (OR=0,64, 95% I.C. 0,43-0,98, p=0,038), resultado que mantiene su significación tras realizar el análisis de estratificación de la muestra acorde al tipo de tumor y al genotipo, comparándolo con la totalidad de muestras control, según: p=0,004, p=0,010 y p=0,006, respectivamente. Sin embargo, atendiendo al grado histológico del tumor, no se evidenció ninguna asociación estadísticamente significativa para ninguno de los SNPs testados.

La consideración del tamaño del tumor relacionó a los SNPs intrónicos rs227060 y rs170548 de *ATM* con tumores de mayor tamaño que 2 cm en su diámetro mayor: OR=1,39, 95% I.C. 1,05-1,85, p=0,022 y OR=1,36, 95% I.C. 1,02-1,80, p=0,035, respectivamente, aunque estos datos no se correlacionaron con riesgo a cáncer de mama por dichos SNPs en una muestra estratificada, al resultar un p-valor no significativo (p=0,243 y p=0,375, respectivamente). Por su parte, los cambios intrónicos rs8176265, en *BRCA1*, y rs3219008 en *OGG1*, han evidenciado asociación a tumores de pequeño tamaño, según: OR=0,73, 95% I.C. 0,54-0,98, p=0,035 (p-valor de estratificación de 0,046) y OR=0,67, 95% I.C. 0,47-0,95, p=0,024 (p-valor de estratificación de 0,013), respectivamente.

c. Análisis de los marcadores tumorales:

El análisis de los marcadores tumorales puso de manifiesto que no hay ninguna relación evidente entre tumores ER y PR positivos y los diferentes SNPs estudiados en este grupo de genes.

IV. RESULTADOS

IV.2. Estudio de asociación a cáncer de mama esporádico de genes implicados en la reparación del ADN

Tabla 13A: Análisis estadístico de los genotipos de los genes relacionados con la familia de genes *FANC* frente al tipo de tumor.

Gen	SNP	Tumor invasivo		Grado de tumor > 1		Tamaño de tumor > 2cm	
		OR (I.C. 95%)	p-valor	OR (I.C. 95%)	p-valor	OR (I.C. 95%)	p-valor
ATM	rs228589	0,80 (0,54-1,20)	0,278	0,95 (0,69-1,31)	0,762	0,84 (0,63-1,10)	0,200
	rs189037	0,74 (0,50-1,10)	0,141	0,86 (0,63-1,17)	0,335	0,77 (0,58-1,01)	0,058
	rs625120	0,76 (0,50-1,14)	0,187	0,95 (0,68-1,32)	0,738	0,83 (0,63-1,10)	0,189
	rs228590	0,76 (0,51-1,14)	0,184	0,88 (0,64-1,22)	0,447	0,81 (0,62-1,08)	0,147
	rs228599	0,75 (0,50-1,12)	0,157	0,88 (0,64-1,23)	0,458	0,85 (0,64-1,12)	0,243
	rs599164	0,73 (0,49-1,10)	0,133	0,91 (0,66-1,27)	0,589	0,90 (0,68-1,18)	0,433
	rs637064	0,73 (0,50-1,15)	0,241	0,90 (0,80-1,52)	0,541	0,78 (0,93-1,60)	0,157
	rs647681	0,76 (0,51-1,12)	0,167	0,92 (0,67-1,26)	0,589	0,87 (0,66-1,15)	0,316
	rs1801516	0,97 (0,55-1,72)	0,916	0,81 (0,52-1,26)	0,346	0,72 (0,47-1,08)	0,112
	rs660429	0,76 (0,50-1,13)	0,179	0,87 (0,63-1,20)	0,401	0,81 (0,61-1,07)	0,134
	rs227060	1,22 (0,79-1,88)	0,381	1,06 (0,76-1,46)	0,748	1,39 (1,05-1,85)	0,022
	rs227064	0,74 (0,49-1,12)	0,154	0,89 (0,65-1,24)	0,503	0,84 (0,63-1,11)	0,218
	rs664143	0,72 (0,48-1,08)	0,107	0,88 (0,64-1,22)	0,442	0,85 (0,64-1,22)	0,248
	rs170548	1,13 (0,74-1,74)	0,566	1,06 (0,76-1,46)	0,748	1,36 (1,02-1,80)	0,035
	rs4585	0,76 (0,50-1,14)	0,180	0,89 (0,64-1,22)	0,460	0,81 (0,61-1,07)	0,142
ATR	rs6782400	0,81 (0,52-1,26)	0,355	0,97 (0,68-1,37)	0,845	0,99 (0,74-1,32)	0,927
	rs7636909	0,93 (0,59-1,46)	0,754	1,01 (0,70-1,44)	0,969	0,97 (0,72-1,30)	0,823
	rs12635931	1,49 (0,72-3,07)	0,279	1,10 (0,66-1,83)	0,726	1,18 (0,79-1,78)	0,423
BRCA1	rs799923	0,84 (0,49-1,44)	0,524	1,38 (0,87-2,21)	0,175	1,03 (0,71-1,49)	0,863
	rs16941	0,62 (0,40-0,96)	0,030	0,98 (0,70-1,37)	0,896	0,75 (0,55-1,01)	0,059
	rs8176265	0,65 (0,43-0,99)	0,046	1,00 (0,72-1,39)	1,000	0,73 (0,54-0,98)	0,035
	rs8176318	0,64 (0,43-0,98)	0,038	1,01 (0,72-1,41)	0,956	0,77 (0,57-1,04)	0,083
H2AX	rs8551	1,15 (0,76-1,73)	0,509	0,81 (0,58-1,12)	0,198	0,91 (0,69-1,20)	0,490
	rs7759	0,95 (0,63-1,45)	0,817	0,79 (0,57-1,11)	0,171	1,03 (0,78-1,37)	0,843
	rs7350	0,93 (0,62-1,41)	0,746	0,85 (0,61-1,18)	0,320	1,05 (0,79-1,39)	0,728
OGG1	rs159153	1,25 (0,80-1,98)	0,332	0,83 (0,59-1,18)	0,307	1,02 (0,75-1,38)	0,893
	rs125701	1,08 (0,59-1,96)	0,815	0,67 (0,43-1,05)	0,080	1,00 (0,68-1,47)	0,995
	rs3219008	1,03 (0,62-1,69)	0,918	1,21 (0,81-1,82)	0,361	0,67 (0,47-0,95)	0,024
	rs1052133	1,19 (0,70-2,05)	0,518	1,30 (0,85-2,01)	0,231	0,79 (0,55-1,12)	0,186
USP1	rs9436223	1,07 (0,68-1,68)	0,782	1,09 (0,76-1,57)	0,631	1,19 (0,88-1,60)	0,255
	rs583609	0,99 (0,63-1,55)	0,969	1,05 (0,73-1,52)	0,779	1,21 (0,90-1,62)	0,215

Análisis estadístico del efecto aditivo del genotipo (en su alelo menor), para los genes relacionados con la familia de genes *FANC* entre los diferentes grupos de muestras segregadas en función de las características tumorales: tumor invasivo *versus* tumor *in situ*, tumor de grado >1 *versus* tumor de grado ≤1 y tumor de tamaño >2 cm *versus* tumor de tamaño ≤2 cm. Los resultados estadísticamente significativos se indican en negrita rojo. Los resultados marginales se indican en negrita negro.

d. Análisis del estado de metástasis:

En cuanto a la relación con la capacidad metastática del tumor, se determinó la asociación del cambio intrónico rs637064 de *ATM* con pacientes sin metástasis a ganglios linfáticos (OR=0,63, 95% I.C. 1,03-1,82, p=0,029), efecto que se ve confirmado tras la estratificación de la muestra acorde a este hecho (p=0,008). El estudio de la metástasis a distancia, por su parte, mostró que el cambio rs125701, localizado a 5' corriente arriba de *OGG1*, está asociado a una situación de metástasis a tejidos distales de la mama, con un riesgo de 2,55 (95% I.C. 1,21-5,37, p=0,014), y cuya asociación se presenta también de un modo significativo en el estudio de asociación a cáncer de mama con estratificación de la muestra por esta característica (p=0,009).

Tabla 13B: Análisis estadístico de los genotipos de los genes relacionados con la familia de genes *FANC* frente al estado de los marcadores tumorales y de metástasis.

Gen	SNP	ER Positivo		PR Positivo		Metástasis a ganglio		Otras metástasis	
		OR (I.C. 95%)	p-valor	OR (I.C. 95%)	p-valor	OR (I.C. 95%)	p-valor	OR (I.C. 95%)	p-valor
ATM	rs228589	1,11 (0,78-1,57)	0,571	1,21 (0,90-1,63)	0,207	0,78 (0,59-1,04)	0,087	0,88 (0,46-1,67)	0,691
	rs189037	1,23 (0,88-1,74)	0,232	1,24 (0,93-1,66)	0,146	0,76 (0,58-1,01)	0,059	1,05 (0,57-1,95)	0,877
	rs625120	1,17 (0,82-1,67)	0,385	1,34 (0,99-1,82)	0,061	0,80 (0,60-1,07)	0,139	1,19 (0,63-2,42)	0,586
	rs228590	1,13 (0,80-1,60)	0,499	1,30 (0,96-1,75)	0,087	0,78 (0,59-1,04)	0,093	1,18 (0,63-2,22)	0,601
	rs228599	1,13 (0,79-1,61)	0,502	1,31 (0,97-1,77)	0,083	0,80 (0,60-1,07)	0,135	1,13 (0,59-2,17)	0,718
	rs599164	1,15 (0,80-1,63)	0,456	1,35 (1,00-1,84)	0,051	0,83 (0,62-1,10)	0,198	1,02 (0,53-1,94)	0,956
	rs637064	1,03 (0,68-1,35)	0,793	1,14 (0,64-1,15)	0,300	0,63 (1,03-1,82)	0,029	1,10 (0,48-1,69)	0,741
	rs647681	1,05 (0,75-1,49)	0,769	1,31 (0,98-1,77)	0,072	0,82 (0,62-1,09)	0,180	1,01 (0,53-1,92)	0,976
	rs1801516	0,75 (0,47-1,21)	0,238	0,87 (0,57-1,33)	0,524	1,08 (0,74-1,59)	0,679	0,40 (0,10-1,61)	0,197
	rs660429	1,14 (0,80-1,61)	0,479	1,31 (0,97-1,76)	0,082	0,77 (0,58-1,02)	0,067	1,19 (0,64-2,23)	0,583
	rs227060	0,90 (0,63-1,30)	0,583	0,82 (0,60-1,12)	0,202	1,07 (0,80-1,42)	0,658	0,93 (0,48-1,82)	0,837
	rs227064	1,12 (0,79-1,60)	0,524	1,31 (0,97-1,78)	0,082	0,78 (0,58-1,05)	0,096	1,04 (0,55-1,98)	0,902
	rs664143	1,14 (0,80-1,63)	0,469	1,33 (0,98-1,81)	0,065	0,79 (0,59-1,06)	0,112	1,03 (0,54-1,96)	0,927
	rs170548	0,92 (0,64-1,32)	0,650	0,84 (0,62-1,15)	0,276	1,06 (0,80-1,42)	0,686	0,94 (0,48-1,83)	0,844
ATR	rs4585	1,12 (0,78-1,60)	0,534	1,34 (0,99-1,82)	0,060	0,75 (0,56-1,01)	0,056	1,18 (0,63-2,23)	0,606
	rs6782400	1,10 (0,74-1,63)	0,633	1,05 (0,76-1,47)	0,761	0,96 (0,72-1,29)	0,793	0,96 (0,48-1,93)	0,910
	rs7636909	0,89 (0,60-1,32)	0,568	0,92 (0,66-1,30)	0,640	0,81 (0,60-1,09)	0,161	1,11 (0,56-2,22)	0,761
BRCA1	rs12635931	0,81 (0,47-1,38)	0,434	0,90 (0,56-1,44)	0,652	1,10 (0,72-1,66)	0,667	0,60 (0,18-1,97)	0,401
	rs799923	1,05 (0,65-1,69)	0,856	1,13 (0,76-1,70)	0,550	1,13 (0,78-1,63)	0,529	0,80 (0,33-1,98)	0,634
	rs16941	0,85 (0,58-1,24)	0,394	0,78 (0,57-1,08)	0,132	1,01 (0,75-1,37)	0,947	0,79 (0,37-1,66)	0,529
	rs8176265	0,81 (0,56-1,17)	0,268	0,74 (0,54-1,02)	0,064	1,02 (0,76-1,37)	0,908	0,92 (0,47-1,81)	0,807
H2AX	rs8176318	0,82 (0,57-1,18)	0,280	0,74 (0,54-1,02)	0,063	1,03 (0,77-1,39)	0,838	0,90 (0,46-1,77)	0,753
	rs8551	1,17 (0,82-1,68)	0,379	1,24 (0,92-1,69)	0,158	0,93 (0,71-1,23)	0,626	0,91 (0,48-1,74)	0,779
	rs7759	0,99 (0,69-1,42)	0,946	1,11 (0,81-1,52)	0,520	1,09 (0,82-1,44)	0,549	1,11 (0,57-2,17)	0,765
OGG1	rs7350	0,87 (0,61-1,24)	0,424	1,01 (0,74-1,37)	0,977	1,06 (0,80-1,40)	0,702	1,20 (0,63-2,28)	0,583
	rs159153	0,97 (0,65-1,44)	0,881	1,11 (0,79-1,56)	0,534	1,32 (0,97-1,79)	0,074	1,60 (0,83-3,08)	0,160
	rs125701	1,52 (0,86-2,66)	0,148	1,47 (0,93-2,32)	0,102	1,01 (0,68-1,49)	0,963	2,55 (1,21-5,37)	0,014
	rs3219008	0,94 (0,61-1,44)	0,770	0,71 (0,49-1,01)	0,058	1,03 (0,74-1,43)	0,860	0,57 (0,23-1,44)	0,238
USP1	rs1052133	0,92 (0,59-1,43)	0,701	0,71 (0,49-1,03)	0,070	1,00 (0,71-1,41)	0,990	0,74 (0,31-1,77)	0,504
	rs9436223	1,08 (0,72-1,60)	0,716	0,96 (0,69-1,34)	0,820	0,92 (0,68-1,25)	0,596	1,26 (0,64-2,44)	0,505
	rs583609	1,10 (0,74-1,64)	0,645	0,96 (0,69-1,35)	0,825	0,94 (0,69-1,27)	0,675	1,08 (0,53-2,18)	0,833

Análisis estadístico del efecto aditivo del genotipo (en su alelo menor) para los genes relacionados con la familia de genes *FANC* entre los diferentes grupos de muestras segregadas en función de las características tumorales: tumor ER positivo *versus* tumor ER negativo, tumor PR positivo *versus* tumor PR negativo, tumor metastatizado a ganglio linfático axilar *versus* tumor no metastatizado a ganglio linfático axilar, y tumor metastatizado a otros tejidos distales *versus* tumor no metastatizado a otros tejidos distales. Los resultados estadísticamente significativos se indican en negrita rojo. Los resultados marginales se indican en negrita negro.

IV.3. ESTUDIO DE ASOCIACIÓN A CÁNCER DE MAMA ESPORÁDICO DE GENES IMPLICADOS EN EL CONTROL DEL CICLO CELULAR

IV.3.1. Genes y SNPs seleccionados de la ruta de control del ciclo celular

Se tomó en consideración para este proyecto un segundo grupo de genes cuyas proteínas estaban implicadas en el control del ciclo celular, al ser éste un factor fundamental en la progresión tumoral, si bien a fecha del comienzo del proyecto no había ningún trabajo que relacionara dicho grupo de genes con susceptibilidad a padecer cáncer de mama de forma directa. Los 15 genes seleccionados para este propósito se exponen en el *Anexo I.12.*, entre los que la familia de genes *E2F*, que codifican para factores de transcripción que regularán la expresión de proteínas que intervienen en el ciclo celular, tiene una especial representación.

Así como ocurría en la selección de SNPs para los genes de reparación del ADN, y siguiendo los criterios expuestos en el apartado III.2.2.1, se procedió a la selección de los 62 SNPs que se presentan, junto con sus características más relevantes, en el *Anexo I.13.*

IV.3.2. Relación de *E2F1* y el cáncer de mama esporádico

IV.3.2.1. Estudio de asociación de *E2F1*

De los cuatro SNPs estudiados en *E2F1*, rs3213180, un cambio localizado en 3'UTR del gen, no sigue el equilibrio de Hardy-Weinberg en los casos de cáncer de mama esporádico, con un p-valor de Pearson de 0,005. La comparación de MAF entre casos y controles sugirió una mayor presencia del alelo menor G de rs3213180 entre casos (OR=1,39, 95% I.C. 0,95-2,02, p=0,087), como se presenta en la *Tabla 14*.

Tabla 14: Estudio de asociación del genotipo de rs3213180 de *E2F1* con el cáncer de mama esporádico.

Modelo estadístico	Genotipo alelos	Estudio N original	
		OR (95% I.C.)	p-valor
Codominante	CG	1,25 (0,83-1,88)	0,288
	GG	3,20 (0,64-15,93)	0,134
Por alelo menor	G-	1,35 (0,94-1,94)	0,102

Relativo a los homocigotos para el alelo mayor “C”, el OR estimado para los homocigotos del alelo raro (“GG”) fue de 3,20 (95% I.C. 0,64-15,93, $p=0,134$).

Sin embargo, tras analizar una población mayor de casos y controles para este SNP hasta llegar a un total de 1.069 CME y 815 CoCM, los resultados confirmaron que los casos no siguen el HWE, y se observó una mayor tendencia hacia el riesgo a padecer cáncer de mama esporádico, con un OR=3,59 (95% I.C. 0,77-16,66, $p=0,081$), si bien seguía sin ser estadísticamente significativa. Sin embargo, la estratificación de la muestra acorde a la edad reveló que cuando las muestras con una edad igual o mayor a 50 años son analizadas de forma exclusiva y comparadas frente al grupo completo de controles de población normal, el riesgo relativo observado para homocigotos raros llega a un valor de 6,00 (95% I.C. 0,71-50,05, $p=0,098$), y la asociación alcanza la significación estadística cuando se estima el valor de riesgo relativo por alelo menor (“G”), según: OR=1,54, 95% I.C. 1,05-2,27, $p=0,027$ (ver *Tabla 15* para más detalle).

Tabla 15: Estudio de asociación del genotipo de rs3213180 de *E2F1* con el cáncer de mama esporádico en una población ampliada y estratificación acorde a la edad al diagnóstico.

Modelo estadístico	Genotipo alelos	Estudio N _{total} =1.884					
		Todos		< 50 años		≥ 50 años	
		OR (95% I.C.)	p-valor	OR (95% I.C.)	p-valor	OR (95% I.C.)	p-valor
Codominante	CG	1,13 (0,82-1,55)	0,451	1,02 (0,61-1,70)	0,945	1,40 (0,91-2,15)	0,123
	GG	3,59 (0,77-16,66)	0,081	1,86 (1,17-20,59)	0,614	6,00 (0,71-50,05)	0,098
Por alelo menor	G-	1,23 (0,92-1,64)	0,157	1,07 (0,66-1,71)	0,791	1,54 (1,05-2,27)	0,027

La población utilizada ha sido: 1.069 CME y 815 CoCM. El corte de edad se ha establecido a los 50 años. Los resultados estadísticamente significativos se indican en **negrita roja**. Los resultados marginales se indican en **negrita negro**.

Los resultados obtenidos del análisis de haplotipos y diplotipos de los dos SNPs estudiados en *E2F1*, rs2071056, localizado en el intrón 4 del gen y rs3213180, en 3’UTR, estimados a través del programa PHASE v2.0, se representan en la *Figura 25A* y la *Figura 25B*, respectivamente. Se tomó el haplotipo mayoritario “TC” (que porta los alelos comunes para ambos SNPs) como referencia y se comparó con las demás combinaciones de haplotipos estimados. Se observó una tendencia del haplotipo “GG” (presente en un 5,3% en la muestra), portador de los alelos raros de ambos SNPs, a estar presente en una mayor frecuencia entre casos que entre controles, aunque sin llegar a la significación estadística (OR=1,33; 95% I.C. 0,91-1,94; $p=0,146$). El diplotipo formado por los homocigotos para el haplotipo mayoritario “TC” fue tomado como grupo de referencia (“TC/TC”) para el análisis estadístico de diplotipos. Se observó un incremento en el riesgo

IV. RESULTADOS

IV.3. Estudio de asociación a cáncer de mama esporádico de genes implicados en el control del ciclo celular

a padecer cáncer de mama ($OR=5,89$; 95% I.C. 0,70-49,24) para aquellos individuos portadores del diplotipo "CG/CG" (presente en un 1,3% en la muestra), aunque no alcanzó la significación estadística ($p=0,102$). Sin embargo, aquellos individuos portadores del diplotipo "CC/CC" (14,0%) con el alelo raro de rs2071056 y el alelo normal de rs3213180, mostraron una tendencia a un efecto de protección ($OR=0,63$; 95% I.C. 0,38-1,02; $p=0,061$).

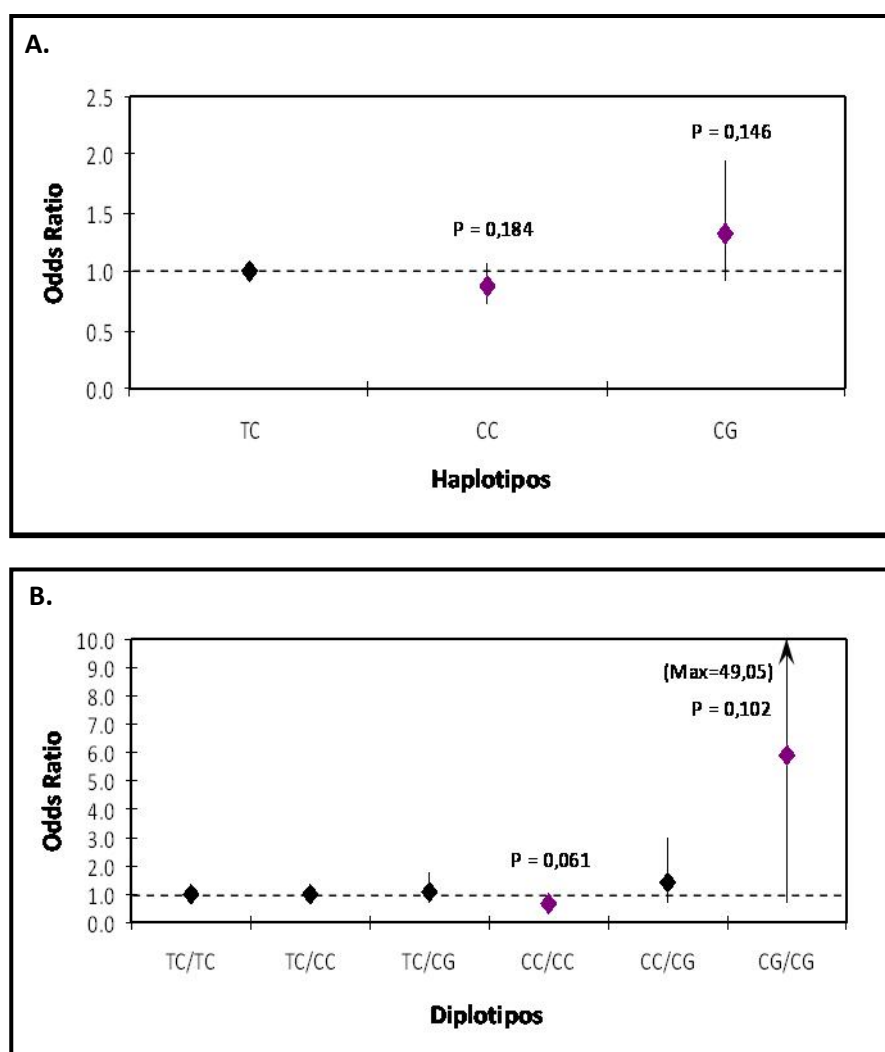


Figura 25: Representación gráfica del análisis de frecuencias haplotípicas (A) y diplotípicas (B) en E2F1. Se indican los ORs (rombos, en morado con estadística marginal) e intervalos de confianza (líneas verticales) del. Los p-valores de interés están indicados. El valor fuera de escala $OR=49,05$ está indicado con una punta de flecha.

IV.3.3.2. Análisis funcional de *E2F1*

El SNP rs3213180 de 3'UTR de *E2F1* induce un incremento de la expresión en ensayos *in vitro*

El SNP rs3213180 está localizado en 3'UTR de *E2F1*, a 29 pb corriente arriba del sitio 2 (S2) de unión de los microRNAs miR-17-20 y miR-20a y alejado corriente abajo del sitio 1 (S1) (O'Donnell *et al.*, 2005). La hipótesis del presente estudio fue determinar si el SNP, debido a su relación con riesgo de padecer cáncer de mama esporádico, está implicado en la des-regulación de los niveles de proteína E2F1 a través de la des-regulación de la unión de los miRNAs en el ARNm de *E2F1*. Para tal propósito, se realizaron diversos experimentos de luciferasa usando células HeLa, que producen de forma constitutiva los miRNAs arriba mencionados.

Se analizó la secuencia completa de 3'UTR de *E2F1*, portando el alelo normal "C", tomado como referencia, o el alelo de riesgo "G" de rs3213180. Los resultados mostraron un incremento en la expresión de proteína estadísticamente significativo cuando el vector de expresión de luciferasa portador del inserto con el alelo "G" fue comparado con el de referencia. Esto se observó en ambos sistemas, con el S2 de unión de miRNAs intacto en las bases de unión de los miRNAs en la "secuencia semilla", así como en la forma mutada, donde se interfiere dicha unión (a partir de ahora, nombrada como MS2), manteniéndose S1 todavía activo, según $p=9,07 \times 10^{-10}$ y $p=8,10 \times 10^{-16}$, respectivamente. Más aún, si bien la construcción portadora del alelo "C" con MS2 presentó una expresión mayor que aquella sin la "secuencia semilla" mutada en S2 ($p=9,32 \times 10^{-14}$), la construcción MS2 portadora del alelo "G" mostró una expresión aún mayor al compararse con su correspondiente forma "C" ($p=0,003$). En todos estos experimentos de transfección y expresión de luciferasa se utilizó un sistema de vector de expresión luciferasa con un inserto de 3'UTR de *PTEN* como control positivo del ensayo, al contener un sitio de unión de los mismos miRNAs testados en el presente estudio. Éste se muta en la "secuencia semilla" de forma similar y se observa el incremento de expresión génica debido al bloque del sistema de miRNAs entre *PTEN*-WS (con el sitio de

unión normal) y *PTEN*-MS (con el sitio de unión mutado), que funcionó correctamente en todos los ensayos: $p=5,93 \times 10^{-8}$.

Añadido a la cuantificación de la expresión de proteína Firefly y la detección de luciferasa, se realizaron experimentos de cuantificación de ARNm que demostraron que las diferencias de expresión de los sistemas portadores de “C” y “G” en 3’UTR de *E2F1* no correspondían a un aumento de la expresión de ARNm ($p=0,835$).

Con el fin de acotar la región de acción del alelo “G”, se evaluó la expresión de luciferasa utilizando dos diferentes fragmentos de 3’UTR incluyendo a rs3213180 y S2, normal o con la “secuencia semilla” mutada, FS2 y FMS2, respectivamente. Los resultados mostraron una mayor expresión, estadísticamente significativa, entre aquellas construcciones portadoras del alelo “G” comparada con la secuencia de referencia cuando S2 está presente ($p=2,75 \times 10^{-13}$). Una expresión similar fue obtenida cuando ambos alelos se testaron en un sistema con MS2, siendo la construcción con el alelo de riesgo “G” inductora de una mayor expresión de luciferasa estadísticamente diferente de aquella con “C” ($p=1,39 \times 10^{-8}$). Finalmente, la exclusión completa de S2 en un sistema portador del SNP con una secuencia corta alrededor, llevó a la total abolición de la expresión de proteína Firefly, y consecuentemente, de *E2F1* en ambos portadores “C” y “G”, para los que no se observó diferencia alguna ($p=0,135$). Los detalles de los ensayos expuestos en este apartado se representan a continuación en la *Figura 26*.

El SNP rs3213180 de 3’UTR de *E2F1* induce cambios del plegamiento del ARNm, evaluados *in silico*

En la *Figura 27* se muestra la predicción de plegamiento del ARNm según la plataforma predictora *GeneBee RNA Secondary Structure Prediction*, acorde a las diferentes construcciones descritas en el apartado anterior. Este análisis *in silico* mostró que tanto la secuencia completa de 3’UTR con rs3213180 y los sitios 1 y 2 de unión de los miRNAs, como el fragmento de 3’UTR que porta el SNP y el sitio 2, cambian estructuralmente cuando “G” es introducido en la secuencia.

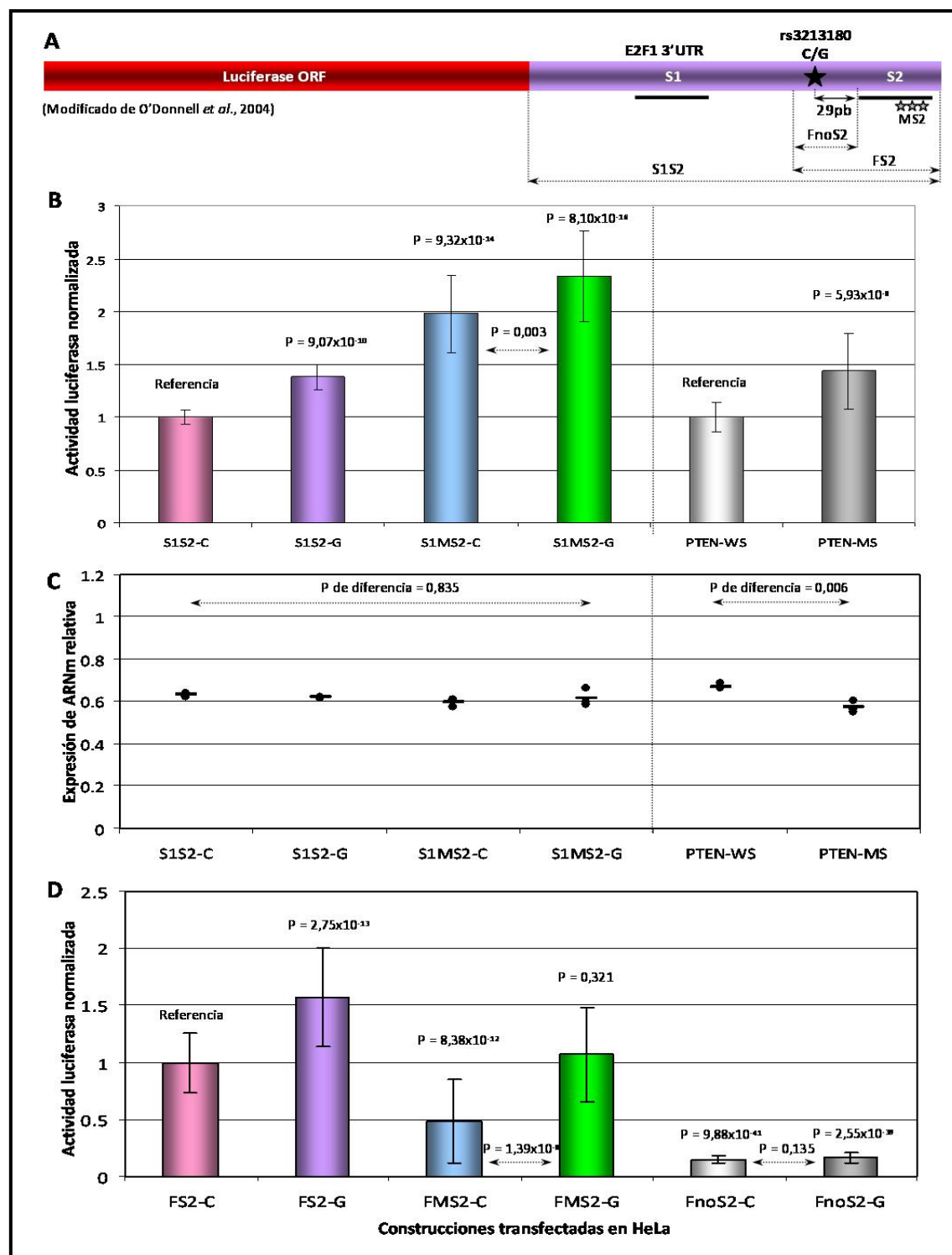


Figura 26: **A**, Representación esquemática de las construcciones utilizadas en los ensayos de luciferasa en células HeLa, indicando la posición de rs3213180, la secuencia 3'UTR portadora de los sitios 1 y 2 de unión de miRNAs (S1S2), el fragmento de 3'UTR que incluye al sitio 2 (FS2), el fragmento que excluye al sitio 2 (FnoS2), y la mutación de la "secuencia semilla" del sitio 2 (MS2, estrellas grises). **B**, Actividad luciferasa normalizada con renilla para las construcciones con 3'UTR completo, portadoras de los alelos "C" y "G". Se muestra también el control del ensayo, *PTEN* normal (WS) y con la "secuencia semilla" mutada (MS). **C**, Valores de expresión de ARNm relativos para cada construcción de B representados como puntos, y sus medias son líneas horizontales. **D**, Actividad luciferasa normalizada con renilla para las construcciones con 3'UTR fragmentado, portadoras de los alelos "C" y "G". La actividad luciferasa se representa en forma de barras, las desviaciones estándar son líneas verticales. Los p-valor de cada ensayo están indicados.

IV. RESULTADOS

IV.3. Estudio de asociación a cáncer de mama esporádico de genes implicados en el control del ciclo celular

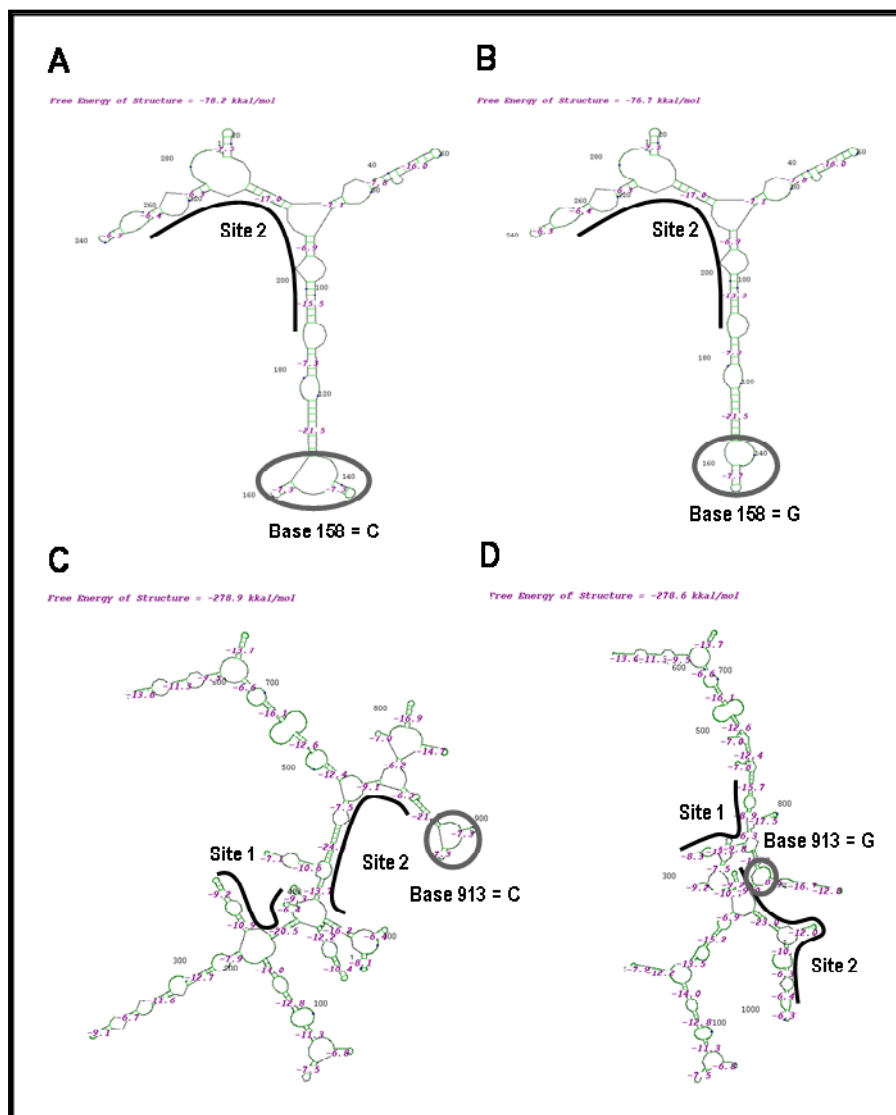


Figura 27: Predicción de plegamiento del ARNm. **A,** Predicción de plegamiento en el fragmento de 3'UTR que incluye a rs3213180 con genotipo "C" y el sitio 2 de unión de miRNAs (FS2). **B,** Predicción de plegamiento en el FS2 con genotipo "G" para rs3213180. **C,** Predicción de plegamiento en la secuencia completa de 3'UTR que incluye a rs3213180 con genotipo "C" y ambos sitios 1 y 2 de unión de miRNAs. **D,** Predicción de plegamiento en la secuencia completa de 3'UTR con genotipo "G" para rs3213180.

El SNP rs3213180 de 3'UTR de *E2F1* induce una mayor expresión *in vivo* de *E2F1* en tejidos de mama no tumorales

A través de ensayos de inmunohistoquímica en tejidos de mama normales embebidos en parafina y provenientes de pacientes genotipados, se llevó a cabo la evaluación de la expresión de *E2F1* nuclear. Se observó una mayor presencia de *E2F1* nuclear entre muestras homocigotos para el alelo "G", aunque no se alcanzó la significación estadística debido al reducido número de muestras recogidas ($p=0,294$).

Además, un análisis de expresión de ARNm de las mismas muestras parafinadas anteriores, reveló unos resultados similares a los obtenidos *in vitro*, al reportar que la tendencia a las diferencias de expresión de E2F1 debidas a rs3213180 no se corresponde con un incremento en la expresión de ARNm (*Figura 28*).

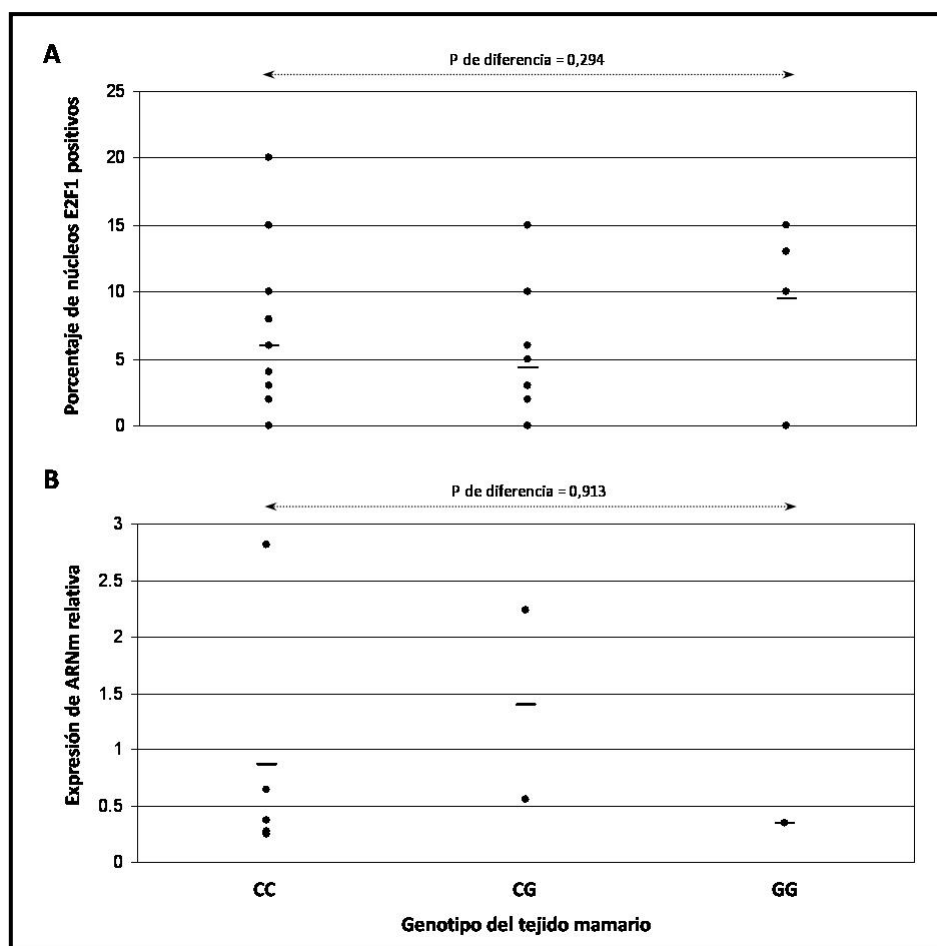


Figura 28: Cuantificación de la expresión de ARNm de *E2F1*. **A**, Porcentaje de núcleos positivos para expresión de E2F1 por cada grupo de genotipo "CC", "CG" y "GG", en tejido mamario normal. **B**, Valores de expresión de ARNm correspondiente a las muestras de A. Los valores están representados como puntos, y sus medias son líneas horizontales. Los p-valores de diferencia entre cada grupo están indicados.

Genotipado adicional de rs3213180 en una población independiente

Se realizó un estudio de replicación de resultados de asociación por parte del cambio rs3213180 en una población caucásica independiente como es la colección de muestras finlandesas de más de 800 casos y controles, del grupo de H. Nevanlinna, del Departamento de Obstetricia y Ginecología de la Universidad de Helsinki. Esta colección es de tamaño similar a la española y fue posible disponer de ella tras un estudio

IV. RESULTADOS

IV.3. Estudio de asociación a cáncer de mama esporádico de genes implicados en el control del ciclo celular

colaborativo, puesto que no era posible alargar nuestro tamaño muestral más allá de las 1.000 muestras por grupo ya genotipadas.

Los resultados de frecuencias alélicas de rs3213180 en el estudio de asociación caso-control de las muestras finlandesas mostró un MAF más parecido al descrito por HapMap (9% y 12%, respectivamente) que el obtenido en las muestras españolas (5%). Sin embargo, no pudimos replicar los datos de asociación obtenidos puesto que el OR para homocigotos para el alelo raro fue de 0,69 (95% I.C. 0,27-1,79, $p=0,442$) (Tabla 16). Se observó que la distribución de las muestras por edad de la colección finlandesa presenta un desbalance de las edades entre casos y controles ($p=2,51 \times 10^{-14}$) (Tabla 17) y el estudio por edad no mostró ninguna asociación a la enfermedad (Tabla 16).

Tabla 16: Distribución de la media de edad en controles para ambas colecciones de muestras, el estudio del Grupo de Genética Humana del CNIO y el grupo de H. Nevanlinna.

Edad de diagnóstico	Estudio N _{total} =1.884			Estudio finlandés		
	Controles (N=815) N (%)	Casos (N=1069) N (%)	p-valor*	Controles (N=915) N (%)	Casos (N=876) N (%)	p-valor*
Rango	23-80	23-89		18-65	22-96	
Mediana (D.E.)	49,9 (11,0)	53,0 (12,7)		41,1 (13,5)	56,5 (12,0)	
< Media	330 (41,2)	360 (41,7)	0,842	616 (67,3)	249 (28,4)	$2,51 \times 10^{-14}$
≥ Media	471 (58,8)	502 (58,3)		299 (32,7)	627 (71,6)	
Desconocido	14 (1,7)	207 (19,4)		0 (0,0)	0 (0,0)	

* Test exacto de Fisher, con p-valor obtenido eliminando los valores desconocidos. D.E., desviación estándar. Los resultados estadísticamente significativos se indican en negrita rojo.

Tabla 17: Estudio de asociación del genotipo de rs3213180 de *E2F1* con el cáncer de mama esporádico en las muestras finlandesas y estratificación acorde a la edad al diagnóstico.

Modelo estadístico	Genotipo alelos	Estudio finlandés					
		Todos		< 50 años		≥ 50 años	
		OR (95% I.C.)	p-valor	OR (95% I.C.)	p-valor	OR (95% I.C.)	p-valor
Codominante	CG	1,07 (0,82-1,38)	0,644	1.18 (0.78-1.79)	0.427	0.91 (0.63-1.32)	0.622
	GG	0,69 (0,27-1,79)	0,442	0.37 (0.05-3.05)	0.357	0.73 (0.20-2.60)	0.622
Por alelo menor	G-	1,01 (0,80-1,27)	0,951	1.05 (0.72-1.53)	0.791	0.90 (0.64-1.25)	0.513

El corte de edad se ha establecido a los 50 años.

IV.3.3. Estudio de asociación de la familia de genes *E2F*

Se llevó a cabo el estudio de los 5 genes *E2F* para los que hubo información de variabilidad genética suficiente como para realizar los análisis de genotipado requeridos: *E2F1*, *E2F2*, *E2F3*, *E2F4* y *E2F5*.

IV.3.3.1. Estudio de asociación de genotipado de los genes de la familia E2F

El análisis de la distribución de los genotipos de los SNPs seleccionados en los genes *E2F* entre CME y CoCM representado en la *Tabla 18*, permitió verificar que todos los SNPs analizados seguían el HWE en CoCM, si bien en CME se encontraron cinco SNPs con tendencia al desequilibrio, y aparte de rs3213180 de *E2F1*, para el que ya se describió en el apartado anterior todos los resultados de asociación de genotipado, el SNP rs2328489 de *E2F3* presenta una desviación estadísticamente significativa de HWE ($p=0,043$).

En la *Tabla 18* se representa el análisis estadístico realizado para la distribución de frecuencias alélicas entre CME y CoCM de los SNPs analizados en los genes *E2F*. En la *Tabla 19* se indica, asimismo, el análisis estadístico de la distribución de genotipos entre CME y CoCM para los genes *E2F*. Para ambos análisis, no se obtuvo ningún resultado estadísticamente significativo.

Tabla 18: Análisis estadístico de HWE, indicación de MAF en cada grupo de muestras y análisis estadístico de la distribución de frecuencias alélicas entre CoCM y CME para la familia de genes *E2F*.

Gen	SNP	CoCM		CME		OR (I.C. 95%)	p-valor
		p-valor HWE	MAF	p-valor HWE	MAF		
E2F1	rs3213150	0,396	0,29	0,513	0,28	0,93 (0,77-1,12)	0,438
	rs2071056	0,375	0,32	0,543	0,30	0,94 (0,78-1,13)	0,498
	rs3213180	0,699	0,05	0,005	0,06	1,39 (0,95-2,02)	0,087
	rs3213183	0,358	0,31	0,836	0,30	0,96 (0,80-1,16)	0,663
E2F2	rs2742976	0,116	0,36	0,145	0,34	0,89 (0,76-1,06)	0,194
	rs3218148	0,544	0,47	0,089	0,47	0,98 (0,83-1,17)	0,822
	rs2075995	0,233	0,50	0,299	0,50	0,99 (0,84-1,17)	0,895
	rs3218211	0,865	0,48	0,069	0,48	0,99 (0,83-1,18)	0,897
E2F3	rs4712494	0,838	0,21	0,735	0,24	1,20 (1,00-1,43)	0,050
	rs911361	1,000	0,42	0,858	0,40	0,92 (0,78-1,37)	0,366
	rs2328489	0,317	0,26	0,043	0,25	0,96 (0,81-1,14)	0,635
	rs4134957	0,183	0,19	1,000	0,19	1,03 (0,85-1,24)	0,792
	rs1322842	0,695	0,32	0,923	0,34	1,10 (0,92-1,32)	0,293
	rs1042317	1,000	0,26	0,920	0,29	1,18 (1,00-1,39)	0,054
E2F4	rs3729639	0,679	0,06	0,086	0,07	1,24 (0,87-1,75)	0,238
E2F5	rs3808538	0,664	0,27	0,824	0,27	0,99 (0,82-1,20)	0,910

Los resultados estadísticamente significativos se indican en negrita rojo. Los resultados marginales se indican en negrita negro.

IV. RESULTADOS

IV.3. Estudio de asociación a cáncer de mama esporádico de genes implicados en el control del ciclo celular

Tabla 19: Análisis estadístico de distribución de genotipos en cada grupo de muestras, CME y CoCM, y de supervivencia entre CME para los genes de la familia *E2F*.

Gen	SNP	Genotipo Alelo aditivo	Univariante		Multivariante		Supervivencia	
			OR (I.C. 95%)	p-valor	OR (I.C. 95%)	p-valor	OR (I.C. 95%)	p-valor
<i>E2F1</i>	rs3213150	C	0,93 (0,77-1,12)	0,445	1,00 (0,79-1,25)	0,964	0,55 (0,27-1,15)	0,112
	rs2071056	T	0,94 (0,79-1,13)	0,505	1,06 (0,85-1,32)	0,594	0,80 (0,44-1,47)	0,476
	rs3213180	C	3,20 (0,94-1,94)	0,102	1,19 (0,79-1,79)	0,398	0,61 (0,16-2,35)	0,474
	rs3213183	C	0,96 (0,80-1,15)	0,667	1,04 (0,83-1,30)	0,758	0,58 (0,29-1,14)	0,111
<i>E2F2</i>	rs2742976	G	0,88 (0,73-1,06)	0,178	1,06 (0,84-1,32)	0,641	0,85 (0,46-1,59)	0,609
	rs3218148	C	0,98 (0,83-1,16)	0,826	0,99 (0,81-1,21)	0,918	1,37 (0,82-2,27)	0,232
	rs2075995	C	0,99 (0,84-1,17)	0,898	1,02 (0,84-1,25)	0,827	0,89 (0,53-1,49)	0,651
	rs3218211	C	0,99 (0,83-1,18)	0,898	0,96 (0,77-1,19)	0,699	1,01 (0,60-1,73)	0,960
<i>E2F3</i>	rs4712494	T	1,30 (1,01-1,43)	0,044	1,17 (0,93-1,47)	0,178	1,13 (0,61-2,08)	0,700
	rs911361	C	0,92 (0,78-1,10)	0,366	0,86 (0,70-1,07)	0,171	1,15 (0,68-1,95)	0,608
	rs2328489	C	0,97 (0,82-1,15)	0,707	1,10 (0,88-1,37)	0,427	1,20 (0,65-2,23)	0,557
	rs4134957	T	1,03 (0,86-1,25)	0,723	1,12 (0,88-1,42)	0,373	1,17 (0,63-2,18)	0,619
	rs1322842	C	1,10 (0,92-1,32)	0,282	1,18 (0,95-1,47)	0,139	0,82 (0,46-1,47)	0,506
	rs1042317	C	1,18 (1,00-1,39)	0,054	1,20 (0,97-1,50)	0,101	0,92 (0,51-1,68)	0,789
<i>E2F4</i>	rs3729639	C	1,22 (0,87-1,72)	0,254	1,11 (0,74-1,67)	0,618	0,66 (0,21-2,07)	0,473
<i>E2F5</i>	rs3808538	A	0,99 (0,82-1,20)	0,911	1,00 (0,79-1,26)	0,989	0,99 (0,53-1,83)	0,962

Se ha aplicado un análisis univariante y multivariante para las variables de confusión: edad de diagnóstico o recogida de muestra, edad de menarquia, número de hijos, estado de menopausia. Los resultados estadísticamente significativos se indican en negrita rojo. Los resultados marginales se indican en negrita negro.

IV.3.3.2. Estudio de asociación de genotipo y fenotipo de los genes *E2F*

a. Análisis de supervivencia:

En un estudio de supervivencia de las pacientes en función de los genotipos para los SNPs seleccionados en los genes *E2F*, no se observó ninguna diferencia de distribución de genotipos para los SNPs analizados.

b. Análisis según el tipo de tumor:

Los resultados del estudio realizado sobre el efecto del genotipo de los SNPs estudiados en los genes *E2F* en las muestras de casos de cáncer de mama segregadas según las diferentes categorías de las características clínicas seleccionadas, se describen en la *Tabla 20A* y la *Tabla 20B*.

Acorde al tipo de tumor, ninguno de los SNPs estudiados ha mostrado una mayor frecuencia entre los tumores invasivos o tumores *in situ*, de modo que esta característica no se relaciona con la familia de genes *E2F*. Sin embargo, el grado histológico del tumor sí se ha visto relacionado con uno de los genes de la familia, *E2F3*. Así pues, el SNP intrónico rs4712494 está asociado a tumores de grado mayor a 1 (OR=1,68, 95% I.C. 1,12-2,53, p=0,010), mientras que otro cambio en un intrón de *E2F3*, rs2328489, se ha relacionado con tumores de bajo grado (OR=0,69, 95% I.C. 0,49-0,98, p=0,038). En ambos casos, dicho resultado se mantuvo en el análisis de riesgo para cáncer de mama de estos

SNPs con la colección de muestras de casos estratificada acorde a esta característica ($p=0,001$ y $p=0,040$, respectivamente). Asimismo, un tamaño del tumor mayor a 2 cm se ha asociado de una manera muy significativa a una frecuencia del alelo menor elevada para el cambio rs4712494 de *E2F3*, con un OR=1,63 (95% I.C. 1,19-2,22, $p<0,001$), que aparece también asociado a cáncer de mama entre las muestras de esta característica ($p<0,001$).

c. Análisis de los marcadores tumorales:

El estudio de los marcadores tumorales no evidenció ninguna relación entre tumores ER y PR positivos y los diferentes SNPs testados.

Tabla 20A: Análisis estadístico de los genotipos de los genes relacionados con la familia de genes *E2F* frente al tipo de tumor.

Gen	SNP	Tumor invasivo		Grado de tumor > 1		Tamaño de tumor > 2cm	
		OR (I.C. 95%)	p-valor	OR (I.C. 95%)	p-valor	OR (I.C. 95%)	p-valor
<i>E2F1</i>	rs3213150	1.38 (0.83-2.27)	0.212	1.03 (0.73-1.47)	0.861	0.88 (0.64-1.21)	0.432
	rs2071056	1.27 (0.80-2.04)	0.315	0.96 (0.68-1.34)	0.791	0.95 (0.71-1.29)	0.750
	rs3213180	1.35 (0.56-3.28)	0.506	0.82 (0.46-1.46)	0.500	0.96 (0.59-1.56)	0.869
	rs3213183	1.22 (0.75-1.98)	0.417	0.94 (0.67-1.33)	0.742	0.89 (0.65-1.21)	0.451
<i>E2F2</i>	rs2742976	0.89 (0.57-1.40)	0.630	1.14 (0.79-1.63)	0.480	0.93 (0.68-1.27)	0.650
	rs3218148	1.03 (0.69-1.53)	0.890	0.99 (0.72-1.37)	0.950	1.21 (0.91-1.59)	0.190
	rs2075995	1.15 (0.77-1.72)	0.492	1.01 (0.73-1.38)	0.965	0.93 (0.71-1.22)	0.586
	rs3218211	0.89 (0.54-1.46)	0.632	1.06 (0.76-1.49)	0.750	1.11 (0.83-1.49)	0.470
<i>E2F3</i>	rs4712494	1.37 (0.83-2.27)	0.220	1.68 (1.12-2.53)	0.010	1.63 (1.19-2.22)	<0,001
	rs911361	0.86 (0.56-1.32)	0.500	0.96 (0.68-1.34)	0.800	1.01 (0.76-1.34)	0.940
	rs2328489	0.82 (0.53-1.26)	0.364	0.69 (0.49-0.98)	0.038	1.04 (0.76-1.41)	0.827
	rs4134957	0.90 (0.54-1.48)	0.668	0.84 (0.57-1.25)	0.393	1.15 (0.82-1.63)	0.420
	rs1322842	1.12 (0.72-1.75)	0.620	1.32 (0.92-1.89)	0.140	1.02 (0.76-1.37)	0.890
	rs1042317	1.42 (0.88-2.30)	0.150	1.24 (0.85-1.79)	0.260	1.01 (0.75-1.37)	0.926
<i>E2F4</i>	rs3729639	0.59 (0.29-1.23)	0.160	0.96 (0.48-1.93)	0.910	1.02 (0.59-1.76)	0.950
<i>E2F5</i>	rs3808538	1.22 (0.74-2.02)	0.440	1.07 (0.74-1.56)	0.720	1.16 (0.84-1.60)	0.360

Análisis estadístico del efecto aditivo del genotipo (en su alelo menor), para los genes relacionados con la familia de genes *E2F* entre los diferentes grupos de muestras segregadas en función de las características tumorales: tumor invasivo *versus* tumor *in situ*, tumor de grado >1 *versus* tumor de grado ≤1 y tumor de tamaño >2 cm *versus* tumor de tamaño ≤2 cm. Los resultados estadísticamente significativos se indican en negrita rojo.

d. Análisis del estado de metástasis:

El análisis estadístico del estado de metástasis del tumor al momento del diagnóstico hacia ganglios linfáticos cercanos a la mama, por su parte, tampoco mostró ninguna relación con los SNPs estudiados en este grupo de genes, mientras que el SNP rs2742976, localizado 5' corriente arriba de *E2F2* (OR=0,40, 95% I.C. 0,17-0,94, $p=0,040$), se ha visto asociado a ausencia de metástasis a tejidos distales de la mama. El análisis de esta variante en el riesgo para cáncer de mama entre las muestras de CME segregadas en función del estado de metástasis ha confirmado la asociación ($p=0,025$).

IV. RESULTADOS

IV.3. Estudio de asociación a cáncer de mama esporádico de genes implicados en el control del ciclo celular

Tabla 20B: Análisis estadístico de los genotipos de los genes relacionados con la familia de genes *E2F* frente al estado de marcadores tumorales y de metástasis.

Gen	SNP	ER Positivo		PR Positivo		Metástasis a ganglio		Otras metástasis	
		OR (I.C. 95%)	p-valor	OR (I.C. 95%)	p-valor	OR (I.C. 95%)	p-valor	OR (I.C. 95%)	p-valor
<i>E2F1</i>	rs3213150	1.11 (0.73-1.69)	0.623	1.14 (0.80-1.61)	0.481	0.74 (0.54-1.03)	0.075	1.06 (0.52-2.14)	0.873
	rs2071056	1.20 (0.80-1.79)	0.381	1.16 (0.83-1.62)	0.382	0.92 (0.68-1.24)	0.577	1.34 (0.69-2.59)	0.393
	rs3213180	1.52 (0.76-3.06)	0.241	1.01 (0.60-1.71)	0.969	0.83 (0.50-1.37)	0.466	0.38 (0.05-2.65)	0.327
	rs3213183	1.22 (0.80-1.84)	0.355	1.17 (0.83-1.65)	0.373	0.84 (0.62-1.15)	0.284	1.23 (0.62-2.44)	0.547
<i>E2F2</i>	rs2742976	0.81 (0.54-1.23)	0.320	0.77 (0.55-1.09)	0.140	1.18 (0.86-1.62)	0.290	0.40 (0.17-0.94)	0.040
	rs3218148	0.79 (0.55-1.12)	0.190	0.82 (0.60-1.10)	0.190	0.91 (0.68-1.20)	0.490	1.02 (0.54-1.94)	0.950
	rs2075995	1.26 (0.90-1.78)	0.202	1.19 (0.90-1.59)	0.250	1.26 (0.96-1.66)	0.098	0.89 (0.47-1.67)	0.719
	rs3218211	0.94 (0.63-1.39)	0.750	0.87 (0.63-1.20)	0.410	0.80 (0.60-1.08)	0.140	1.24 (0.66-2.35)	0.510
<i>E2F3</i>	rs4712494	1.02 (0.67-1.56)	0.920	0.86 (0.60-1.22)	0.390	1.34 (0.97-1.83)	0.070	1.50 (0.76-2.97)	0.240
	rs911361	0.86 (0.59-1.25)	0.430	1.21 (0.88-1.67)	0.250	0.99 (0.75-1.32)	0.950	0.87 (0.44-1.70)	0.680
	rs2328489	1.01 (0.67-1.52)	0.959	1.13 (0.81-1.59)	0.474	1.13 (0.82-1.54)	0.456	1.54 (0.78-3.01)	0.211
	rs4134957	1.02 (0.65-1.61)	0.920	0.94 (0.65-1.37)	0.740	1.27 (0.90-1.80)	0.180	0.93 (0.41-2.15)	0.870
	rs1322842	1.03 (0.69-1.54)	0.870	0.98 (0.70-1.36)	0.900	0.98 (0.73-1.32)	0.900	1.04 (0.52-2.05)	0.920
	rs1042317	1.18 (0.78-1.78)	0.423	0.99 (0.71-1.38)	0.950	1.06 (0.78-1.44)	0.711	1.34 (0.68-2.65)	0.396
<i>E2F4</i>	rs3729639	0.83 (0.41-1.69)	0.610	1.33 (0.71-2.48)	0.380	1.03 (0.59-1.79)	0.920	-	-
<i>E2F5</i>	rs3808538	1.34 (0.86-2.61)	0.190	0.96 (0.68-1.37)	0.840	1.18 (0.85-1.63)	0.320	0.51 (0.21-1.23)	0.130

Análisis estadístico del efecto aditivo del genotipo (en su alelo menor) para los genes relacionados con la familia de genes *E2F* entre los diferentes grupos de muestras segregadas en función de las características tumorales: tumor ER positivo *versus* tumor ER negativo, tumor PR positivo *versus* tumor PR negativo, tumor metastatizado a ganglio linfático axilar *versus* tumor no metastatizado a ganglio linfático axilar, y tumor metastatizado a otros tejidos distales *versus* tumor no metastatizado a otros tejidos distales. Los resultados estadísticamente significativos se indican en negrita rojo. Los resultados marginales se indican en negrita negro.

IV.3.4. Estudio de otros genes relacionados con el control del ciclo celular

Se realizó el estudio de 9 genes que participan en el control del ciclo celular, muchos de ellos relacionados, asimismo, con la familia *E2F*, y son: *BRG1*, *BUB1B*, *c-MYC*, *EP300*, *MDM2*, *RB1*, *RBL1/p130*, *STK6*, *VDR*.

IV.3.4.1. Estudio de asociación de genotipado de otros genes de control del ciclo celular

Se realizó el estudio de HWE para los genotipos de los SNPs seleccionados en los genes no pertenecientes a la familia *E2F* y relacionados con el control del ciclo celular, lo cual permitió descartar del estudio el único SNP estudiado en el gen *BRG1*, debido a su carácter monomórfico y a la desviación del HWE en controles y casos. Otros SNPs han mostrado una desviación marginal, que no ha sido tomada en consideración puesto que se han incluido con normalidad en el análisis posterior (*Tabla 21*).

Tabla 21: Análisis estadístico de HWE, indicación de MAF en cada grupo de muestras y análisis estadístico de la distribución de frecuencias alélicas entre CoCM y CME para otros genes de control del ciclo celular.

Gen	SNP	CoCM		CME		OR (I.C. 95%)	p-valor
		p-valor HWE	MAF	p-valor HWE	MAF		
BRG1	rs1801514	0,000	0	0,000	0	1,03 (0,02-51,86)	1,000
	rs3759839	0,141	0,49	0,169	0,52	1,13 (0,95-1,34)	0,167
	rs3214012	0,610	0,31	0,840	0,32	1,04 (0,86-1,24)	0,702
BUB1B	rs1801376	1,000	0,32	0,753	0,31	0,96 (0,80-1,16)	0,676
	rs2305653	0,380	0,47	0,130	0,44	0,90 (0,76-1,07)	0,213
	rs4924437	0,355	0,30	0,920	0,31	1,03 (0,86-1,24)	0,750
c-MYC	rs2070583	1,000	0,02	1,000	0,02	1,09 (0,61-1,97)	0,767
EP300	rs20552	0,781	0,38	0,565	0,36	0,94 (0,79-1,12)	0,502
MDM2	rs1144943	0,782	0,38	0,784	0,38	1,03 (0,86-1,22)	0,765
	rs937282	0,475	0,47	0,196	0,48	1,05 (0,89-1,24)	0,553
	rs937283	0,795	0,36	0,568	0,36	1,01 (0,85-1,19)	0,945
	rs1470383	0,151	0,16	1,000	0,16	0,99 (0,79-1,25)	0,939
	rs1201644	0,386	0,47	0,602	0,48	1,04 (0,88-1,24)	0,637
	rs3730590	0,447	0,36	0,510	0,36	1,03 (0,86-1,23)	0,785
	rs1695144	1,000	0,01	1,000	0,02	1-58 (0,76-3,28)	0,213
	rs3730646	0,931	0,35	0,645	0,36	1,05 (0,88-1,24)	0,601
	rs3730654	1,000	0,00	1,000	0,00	0,30 (0,03-2,65)	0,395
RB1	rs9568029	0,524	0,29	1,000	0,29	1,04 (0,86-1,27)	0,667
	rs1573601	0,733	0,07	0,500	0,07	0,97 (0,69-1,36)	0,855
	rs2804094	0,703	0,21	0,798	0,21	1,01 (0,82-1,24)	0,929
	rs4151441	0,323	0,03	1,000	0,02	0,58 (0,32-1,06)	0,071
	rs520342	0,441	0,21	1,000	0,21	0,99 (0,81-1,22)	0,958
	rs4151474	1,000	0,02	1,000	0,02	0,65 (0,35-1,20)	0,160
	rs399413	0,389	0,28	0,833	0,28	1,02 (0,85-1,23)	0,824
	rs9568036	0,473	0,42	0,447	0,43	1,04 (0,87-1,25)	0,642
	rs1894255	0,899	0,22	0,800	0,22	1,02 (0,83-1,25)	0,887
	rs1951775	0,400	0,29	0,753	0,29	1,00 (0,83-1,21)	0,979
	rs198575	0,724	0,07	0,160	0,07	1,00 (0,72-1,41)	0,987
	rs198577	0,732	0,07	0,155	0,06	0,94 (0,67-1,32)	0,738
	rs198580	0,236	0,04	0,387	0,05	1,23 (0,82-1,83)	0,324
	rs198590	0,901	0,22	0,710	0,22	1,00 (0,82-1,23)	0,996
	rs367438	0,711	0,06	0,257	0,06	1,06 (0,74-1,51)	0,762
RBL2/p130	rs10748	0,792	0,41	0,724	0,42	1,06 (0,89-1,26)	0,513
STK6	rs1468056	0,401	0,37	1,000	0,39	1,08 (0,90-1,28)	0,408
	rs2273535	0,651	0,17	0,057	0,20	1,21 (0,97-1,50)	0,089
	rs1047972	0,783	0,20	0,881	0,18	0,87 (0,70-1,08)	0,209
	rs6014711	0,488	0,19	0,877	0,17	0,86 (0,69-1,07)	0,175
	rs911160	0,808	0,23	0,652	0,26	1,15 (0,94-1,40)	0,164
	rs1044377	0,742	0,26	0,132	0,30	1,20 (1,00-1,44)	0,056
VDR	rs4516035	0,178	0,39	0,859	0,41	1,10 (0,93-1,31)	0,273
	rs10735810	0,703	0,34	1,000	0,37	1,17 (0,98-1,40)	0,080
	rs731236	0,067	0,43	0,080	0,38	0,82 (0,69-0,98)	0,028
	rs739837	0,438	0,46	0,154	0,46	1,01 (0,85-1,20)	0,899

Los resultados estadísticamente significativos se indican en negrita rojo. Los resultados marginales se indican en negrita negro.

Tan sólo un SNP, rs731236, localizado en el exón 11 de *VDR* y que lleva al cambio sinónimo p.Ile352Ile, mostró una asociación a protección para el cáncer de mama teniendo en cuenta la distribución alélica diferencial entre CME y CoCM (OR=0,82. 95% I.C. 0,69-0,98, p=0,028). Sin embargo, otro SNP de *VDR*, rs10735810, esta vez un cambio

no sinónimo (p.Met1Thr), presentó una tendencia a estar asociado a riesgo para la enfermedad, con un OR=1,17 y un p-valor marginal de 0,080 (95% I.C. 0,98-1,40).

Como ha ocurrido en estudios anteriores, se ha observado una correlación de los resultados obtenidos para las frecuencias alélicas y los detectados en el estudio de la distribución de las frecuencias de genotipos y su relación con el cáncer de mama esporádico, como se indica en la *Tabla 22*. De este modo, se detectó una asociación a protección frente a la enfermedad estadísticamente significativa por parte del cambio sinónimo rs731236 de *VDR* (OR=0,84, 95% I.C. 0,71-0,99, p=0,034), cuya significación se pierde en el análisis multivariante (p=0,102), no así su tendencia a protección. Asimismo, el SNP no sinónimo rs10735810 de *VDR* presenta asociación a riesgo marginal (OR=1,17, 95% I.C. 0,98-1,40, p=0,081), que alcanza la significación estadística (p=0,036) una vez se consideran los factores de confusión anteriormente citados.

IV.3.4.2. Estudio de asociación de genotipo y fenotipo de otros genes de control del ciclo celular

a. Análisis de supervivencia:

Muchos son los SNPs que aparecen asociados a un mejor o peor pronóstico para la evolución del cáncer de mama, fundamentalmente en *BUB1B*. De los cinco cambios estudiados, tres de ellos, rs3214012 (intrónico), rs1801376 (codificante no sinónimo, p.Arg349Gln) y rs2305653 (intrónico), están asociados con una peor supervivencia a la enfermedad, con p-valores inferiores a 0,001. Sin embargo, el SNP seleccionado en la región 5' corriente arriba del gen, rs3759839, está asociado a una mejor supervivencia (OR=0,46, 95% I.C. 0,26-0,80, p=0,006). En el gen *MDM2*, por su parte, encontramos un SNP de carácter intrónico que está asociado a riesgo de fallecimiento por la enfermedad, si bien el amplio intervalo de confianza que presenta debido a su baja MAF, desestima que la asociación sea real (OR=23,63, 95% I.C. 3,09-189,74, p=0,002) (ver *Tabla 22*).

Tabla 22: Análisis estadístico de distribución de genotipos en cada grupo de muestras, CME y CoCM, y de supervivencia entre CME para otros genes de control del ciclo celular.

Gen	SNP	Genotipo Alelo aditivo	Univariante		Multivariante		Supervivencia	
			OR (I.C. 95%)	p-valor	OR (I.C. 95%)	p-valor	OR (I.C. 95%)	p-valor
BUB1B	rs3759839	G	1,13 (0,96-1,34)	0,154	1,12 (0,91-1,38)	0,274	0,46 (0,26-0,80)	0,006
	rs3214012	G	1,03 (0,86-1,24)	0,737	1,08 (0,86-1,35)	0,506	2,54 (1,55-4,15)	0,000
	rs1801376	A	0,96 (0,80-1,16)	0,674	0,98 (0,78-1,22)	0,830	2,44 (1,48-4,02)	0,000
	rs2305653	T	0,89 (0,75-1,06)	0,194	0,87 (0,70-1,07)	0,188	2,51 (1,43-4,40)	0,001
	rs4924437	A	1,03 (0,85-1,23)	0,788	0,98 (0,78-1,22)	0,836	1,48 (0,88-2,51)	0,140
c-MYC	rs2070583	A	1,10 (0,60-1,99)	0,765	1,91 (0,91-3,97)	0,086	0,59 (0,08-4,33)	0,600
EP300	rs20552	T	0,94 (0,79-1,13)	0,501	0,89 (0,72-1,11)	0,298	1,49 (0,82-2,72)	0,189
MDM2	rs1144943	T	1,03 (0,87-1,23)	0,721	0,95 (0,77-1,18)	0,654	0,94 (0,54-1,63)	0,812
	rs937282	C	1,05 (0,90-1,23)	0,544	0,95 (0,78-1,16)	0,610	1,11 (0,66-1,85)	0,697
	rs937283	A	1,01 (0,85-1,21)	0,886	0,92 (0,74-1,15)	0,458	0,99 (0,56-1,77)	0,985
	rs1470383	C	0,99 (0,79-1,24)	0,925	1,03 (0,78-1,35)	0,840	1,04 (0,53-2,03)	0,921
	rs1201644	T	1,05 (0,88-1,24)	0,609	0,94 (0,77-1,16)	0,574	1,03 (0,61-1,75)	0,905
	rs3730590	T	1,03 (0,86-1,23)	0,776	0,91 (0,73-1,13)	0,394	0,90 (0,50-1,61)	0,711
	rs1695144	G	1,59 (0,76-3,31)	0,215	2,24 (0,91-5,46)	0,078	0,66 (0,09-4,90)	0,684
	rs3730646	G	1,05 (0,89-1,25)	0,546	0,92 (0,40-1,15)	0,474	0,95 (0,54-1,67)	0,852
	rs3730654	C	0,29 (0,03-2,60)	0,269	0,52 (0,05-5,15)	0,578	23,63 (3,09-180,74)	0,002
RB1	rs9568029	C	1,05 (0,86-1,27)	0,650	1,21 (0,94-1,55)	0,140	2,00 (0,95-4,21)	0,070
	rs1573601	C	1,01 (0,82-1,24)	0,928	1,06 (0,82-1,37)	0,657	1,75 (0,99-3,09)	0,054
	rs2804094	G	0,97 (0,69-1,36)	0,844	1,09 (0,72-1,64)	0,698	1,22 (0,48-3,09)	0,683
	rs4151441	T	0,58 (0,32-1,06)	0,077	0,77 (0,36-1,64)	0,494	0,05 (0,00-228,16)	0,478
	rs520342	C	1,00 (0,81-1,23)	0,982	1,04 (0,80-1,35)	0,775	1,67 (0,94-2,97)	0,079
	rs4151474	T	0,64 (0,34-1,19)	0,157	0,85 (0,39-1,84)	0,676	0,05 (0,00-297,33)	0,492
	rs399413	G	1,02 (0,85-1,24)	0,805	1,10 (0,87-1,39)	0,428	1,70 (0,99-2,91)	0,055
	rs9568036	A	1,04 (0,87-1,25)	0,637	1,18 (0,94-1,47)	0,154	1,21 (0,60-2,47)	0,593
	rs1894255	A	1,02 (0,83-1,25)	0,863	1,06 (0,82-1,37)	0,670	1,52 (0,85-2,71)	0,161
	rs1951775	G	1,01 (0,83-1,21)	0,963	1,10 (0,87-1,39)	0,434	1,59 (0,93-2,72)	0,094
	rs198575	T	1,00 (0,71-1,41)	0,996	1,13 (0,74-1,71)	0,578	1,27 (0,48-3,34)	0,630
	rs198577	C	0,94 (0,67-1,33)	0,726	1,07 (0,71-1,62)	0,757	1,32 (0,50-3,47)	0,578
	rs198580	A	1,22 (0,82-1,84)	0,328	1,25 (0,77-2,03)	0,377	1,66 (0,63-4,38)	0,306
	rs198590	G	1,00 (0,82-1,23)	0,981	1,03 (0,80-1,33)	0,815	1,42 (0,79-2,55)	0,247
	rs367438	T	1,06 (0,74-1,51)	0,766	1,14 (0,74-1,75)	0,564	1,37 (0,52-3,62)	0,521
RBL2/p130	rs10748	T	1,06 (0,89-1,26)	0,513	1,12 (0,91-1,38)	0,286	1,00 (0,59-1,71)	1,000
STK6	rs1468056	G	1,08 (0,90-1,29)	0,403	1,23 (0,99-1,53)	0,058	1,36 (0,80-2,30)	0,253
	rs2273535	A	1,22 (0,98-1,53)	0,081	1,06 (0,81-1,39)	0,671	1,09 (0,54-2,21)	0,808
	rs1047972	G	0,87 (0,70-1,08)	0,208	0,85 (0,65-1,10)	0,217	0,79 (0,25-1,79)	0,567
	rs6014711	G	0,86 (0,68-1,07)	0,172	0,82 (0,62-1,08)	0,158	0,42 (0,14-1,18)	0,099
	rs911160	G	1,15 (0,95-1,41)	0,161	1,02 (0,80-1,30)	0,881	0,79 (0,40-1,56)	0,494
	rs1044377	G	1,19 (0,99-1,43)	0,060	1,08 (0,86-1,34)	0,508	0,67 (0,36-1,27)	0,224
VDR	rs4516035	T	1,09 (0,92-1,29)	0,304	1,01 (0,82-1,25)	0,905	1,18 (0,68-2,04)	0,555
	rs10735810	G	1,17 (0,98-1,40)	0,081	1,26 (1,02-1,57)	0,036	1,04 (0,62-1,76)	0,874
	rs731236	T	0,84 (0,71-0,99)	0,034	0,84 (0,69-1,03)	0,102	1,19 (0,69-2,06)	0,528
	rs739837	T	1,01 (0,86-1,20)	0,869	1,12 (0,92-1,38)	0,246	1,07 (0,62-1,85)	0,808

Se ha aplicado un análisis univariante y multivariante para las variables de confusión: edad de diagnóstico o recogida de muestra, edad de menarquia, número de hijos, estado de menopausia. Los resultados estadísticamente significativos se indican en negrita rojo. Los resultados marginales se indican en negrita negro.

b. Análisis según el tipo de tumor:

Los resultados que se presentan en la *Tabla 23A* y la *Tabla 23B*, hacen referencia a los análisis realizados con el fin de detectar nuevas asociaciones de los SNPs estudiados para este grupo de genes, atendiendo a las características clínicas indicadas previamente.

El grupo de pacientes con tumores *in situ* presentan una mayor frecuencia para los SNPs rs1047972 (cambio no sinónimo, p.Val57Ile) y rs6014711 (intrónico) de *STK6*, según: OR=0,50, 95% I.C. 0,31-0,82, p=0,006 y OR=0,52, 95% I.C. 0,31-0,86, p=0,010, respectivamente. Ambos SNPs se han relacionado con el cáncer de mama esporádico al estratificar la muestra acorde al tipo de tumor, estableciendo de nuevo la asociación entre los diagnosticados con tumor *in situ* (p=0,018 y p=0,017, respectivamente). En cuanto al grado histológico del tumor, no se han detectado asociaciones de mención con ninguno de los SNPs estudiados.

Al igual que ocurrió con la asociación a una peor supervivencia, un tamaño de tumor mayor de 2 cm está claramente asociado a los SNPs rs3214012 (OR=1,65, 95% I.C. 1,22-2,25, p=0,001), rs1801376 (OR=1,62, 95% I.C. 1,17-2,24, p=0,004), rs2305653 (OR=1,40, 95% I.C. 1,03-1,89, p=0,030) y rs4924437 (OR=1,59, 95% I.C. 1,16-2,18, p=0,004) de *BUB1B*. Todos han evidenciado asociación a cáncer de mama una vez se estratifica la muestra acorde a esta característica, con p-valores de: 0,018, 0,046, 0,012 y 0,042, respectivamente. Asimismo, rs3759839, que se asociaba con un mejor pronóstico, se ve de igual modo asociado a tumores de menor tamaño (OR=0,70, 95% I.C. 0,52-0,95, p=0,020) y a la enfermedad en este grupo de muestras (p=0,006). El SNP intrónico rs9568036 de *RB1*, está asociado, por su parte, a tumores de menor tamaño (OR=0,73, 95% I.C. 0,54-1,00, p=0,047), aunque dicha asociación no implica asociación a cáncer de mama de este SNP para este grupo de muestras (p=0,181).

c. Análisis de los marcadores tumorales:

El estudio de asociación al marcador tumoral ER determinó la amplia asociación de los SNPs de *RB1* a tumores ER positivos, aunque tan sólo dos de ellos, rs1573601, localizado a 5' corriente arriba del gen, y rs399413, un SNP intrónico, con asociación a los tumores ER positivos de OR=2,00 (95% I.C. 1,19-3,40, p=0,010) y OR=1,76 (95% I.C. 1,12-2,77, p=0,015), respectivamente, se asociaron a riesgo para cáncer de mama entre los pacientes con tumores ER positivos, con p-valores de 0,036 y 0,048, respectivamente. Para PR, sin embargo, no se obtuvo asociación alguna.

d. Análisis del estado de metástasis:

El estudio de la distribución de genotipos en función de la metástasis a ganglios linfáticos mostró un número elevado de resultados de asociación marginales, y tres

Tabla 23A: Análisis estadístico de los genotipos de los genes relacionados con otros genes de control del ciclo celular frente al tipo de tumor.

Gen	SNP	Tumor invasivo		Grado de tumor > 1		Tamaño de tumor > 2cm	
		OR (I.C. 95%)	p-valor	OR (I.C. 95%)	p-valor	OR (I.C. 95%)	p-valor
BUB1B	rs3759839	0,90 (0,59-1,39)	0,641	0,90 (0,65-1,26)	0,549	0,70 (0,52-0,95)	0,020
	rs3214012	1,29 (0,81-2,06)	0,289	1,39 (0,96-2,01)	0,082	1,65 (1,22-2,25)	0,001
	rs1801376	1,24 (0,77-2,01)	0,382	1,33 (0,90-1,97)	0,150	1,62 (1,17-2,24)	0,004
	rs2305653	1,14 (0,73-1,79)	0,573	1,19 (0,84-1,68)	0,339	1,40 (1,03-1,89)	0,030
	rs4924437	1,36 (0,84-2,19)	0,212	1,05 (0,73-1,51)	0,777	1,59 (1,16-2,18)	0,004
c-MYC	rs2070583	0,77 (0,22-2,73)	0,684	1,66 (0,43-5,97)	0,435	2,02 (0,84-4,88)	0,118
EP300	rs20552	0,95 (0,61-1,50)	0,838	0,84 (0,58-1,21)	0,342	1,04 (0,75-1,43)	0,833
MDM2	rs1144943	0,95 (0,61-1,46)	0,806	1,01 (0,71-1,43)	0,960	1,12 (0,84-1,49)	0,445
	rs937282	0,95 (0,62-1,44)	0,903	1,22 (0,87-1,70)	0,252	1,13 (0,86-1,48)	0,395
	rs937283	0,96 (0,61-1,50)	0,947	0,92 (0,64-1,33)	0,760	1,09 (0,81-1,46)	0,591
	rs1470383	0,69 (0,41-1,16)	0,160	0,96 (0,61-1,52)	0,865	0,85 (0,57-1,25)	0,406
	rs1201644	1,04 (0,68-1,59)	0,869	1,22 (0,87-1,70)	0,257	1,14 (0,86-1,50)	0,359
	rs3730590	0,97 (0,62-1,52)	0,889	1,02 (0,71-1,46)	0,919	1,02 (0,75-1,37)	0,918
	rs1695144	0,59 (0,16-2,16)	0,430	0,96 (0,29-3,13)	0,946	1,38 (0,52-3,62)	0,520
	rs3730646	1,01 (0,65-1,57)	0,976	0,96 (0,68-1,37)	0,833	1,07 (0,80-1,43)	0,649
	rs3730654	-	-	-	-	-	-
	rs9568029	0,97 (0,62-1,52)	0,889	1,09 (0,74-1,61)	0,647	0,90 (0,64-1,25)	0,518
RB1	rs1573601	0,80 (0,49-1,32)	0,389	1,08 (0,72-1,61)	0,727	0,96 (0,68-1,35)	0,820
	rs2804094	1,21 (0,50-2,93)	0,680	1,06 (0,53-2,11)	0,863	0,70 (0,38-1,29)	0,247
	rs4151441	2,02 (0,26-15,68)	0,502	1,75 (0,37-8,23)	0,481	1,06 (0,36-3,16)	0,915
	rs520342	0,81 (0,50-1,33)	0,404	1,10 (0,73-1,66)	0,641	0,97 (0,69-1,35)	0,838
	rs4151474	2,02 (0,26-15,72)	0,501	1,01 (0,26-3,87)	0,994	0,76 (0,24-2,43)	0,645
	rs399413	0,93 (0,59-1,46)	0,751	1,08 (0,75-1,55)	0,692	0,89 (0,65-1,21)	0,443
	rs9568036	0,88 (0,58-1,34)	0,551	1,06 (0,75-1,51)	0,739	0,73 (0,54-1,00)	0,047
	rs1894255	0,87 (0,53-1,43)	0,573	1,14 (0,76-1,71)	0,539	0,94 (0,67-1,31)	0,706
	rs1951775	0,90 (0,57-1,41)	0,651	1,07 (0,74-1,53)	0,736	0,82 (0,60-1,12)	0,208
	rs198575	1,22 (0,50-3,01)	0,663	0,97 (0,49-1,93)	0,928	0,70 (0,37-1,30)	0,257
	rs198577	1,11 (0,45-2,76)	0,815	0,99 (0,49-2,01)	0,974	0,77 (0,41-1,45)	0,417
	rs198580	0,99 (0,37-2,64)	0,976	0,90 (0,41-1,96)	0,793	0,56 (0,26-1,20)	0,135
	rs198590	0,80 (0,49-1,31)	0,378	1,18 (0,78-1,79)	0,424	0,87 (0,62-1,22)	0,428
	rs367438	1,05 (0,42-2,60)	0,915	1,07 (0,52-2,22)	0,852	0,72 (0,38-1,37)	0,320
RBL2/p130	rs10748	1,10 (0,72-1,69)	0,664	0,88 (0,63-1,23)	0,451	1,10 (0,83-1,46)	0,525
STK6	rs1468056	1,10 (0,72-1,68)	0,650	0,86 (0,63-1,19)	0,370	0,77 (0,57-1,03)	0,073
	rs2273535	1,24 (0,70-2,20)	0,454	1,44 (0,92-2,28)	0,114	0,95 (0,65-1,37)	0,767
	rs1047972	0,50 (0,31-0,82)	0,006	0,77 (0,50-1,18)	0,228	0,76 (0,52-1,13)	0,175
	rs6014711	0,52 (0,31-0,86)	0,010	0,89 (0,57-1,39)	0,610	0,70 (0,47-1,05)	0,088
	rs911160	1,45 (0,86-2,45)	0,162	1,24 (0,85-1,82)	0,263	1,05 (0,76-1,45)	0,773
	rs1044377	1,06 (0,68-1,65)	0,802	1,04 (0,74-1,45)	0,835	1,06 (0,79-1,41)	0,713
VDR	rs4516035	0,87 (0,57-1,32)	0,509	0,93 (0,67-1,30)	0,676	1,02 (0,77-1,35)	0,919
	rs10735810	0,83 (0,54-1,28)	0,396	1,19 (0,84-1,68)	0,332	0,98 (0,73-1,30)	0,875
	rs731236	1,10 (0,72-1,69)	0,663	1,22 (0,87-1,70)	0,257	1,14 (0,86-1,51)	0,367
	rs739837	0,75 (0,49-1,15)	0,192	0,76 (0,54-1,05)	0,094	0,81 (0,61-1,08)	0,151

Análisis estadístico del efecto aditivo del genotipo (en su alelo menor), para otros del ciclo celular, entre los diferentes grupos de muestras segregadas en función de las características tumorales: tumor invasivo *versus* tumor *in situ*, tumor de grado >1 *versus* tumor de grado ≤1 y tumor de tamaño >2 cm *versus* tumor de tamaño ≤2 cm. Los resultados estadísticamente significativos se indican en negrita rojo.

estadísticamente significativos. Dos SNPs de éstos últimos se encuentran en *BUB1B*, aunque presentan sentidos contrapuestos, según: rs3759839 se asocia con ausencia de metástasis a ganglios linfáticos (OR=0,70, 95% I.C. 0,52-0,94, p=0,018), y muestra asociación con cáncer de mama del mismo modo (p=0,006), mientras que rs2305653 es más frecuente entre las muestras con metástasis a ganglios linfáticos (OR=1,46, 95% I.C. 1,07-1,99, p=0,016) y, de igual modo, se asocia con la enfermedad atendiendo a este grupo de pacientes (p=0,008). Este SNP, además, se presenta de una forma más

IV. RESULTADOS

IV.3. Estudio de asociación a cáncer de mama esporádico de genes implicados en el control del ciclo celular

frecuente entre los pacientes con metástasis distal (OR=2,30, 95% I.C. 1,12-4,73, p=0,023), aunque no se asocia con el cáncer de mama estratificando la muestra de acuerdo a esta característica. El cambio no sinónimo de *STK6*, rs1047972 (p.Val57Ile), por su parte, tiende a acumularse entre los pacientes con metástasis nodal (OR=1,53, 95% I.C. 1,06-2,23, p=0,025) y a asociarse con el cáncer de mama en estos casos (p=0,011).

Tabla 23B: Análisis estadístico de los genotipos de otros genes de control del ciclo celular frente al estado de marcadores tumorales y de metástasis.

Gen	SNP	ER Positivo		PR Positivo		Metástasis a ganglio		Otras metástasis	
		OR (I.C. 95%)	p-valor	OR (I.C. 95%)	p-valor	OR (I.C. 95%)	p-valor	OR (I.C. 95%)	p-valor
<i>BUB1B</i>	rs3759839	0,95 (0,65-1,40)	0,807	0,86 (0,62-1,18)	0,340	0,70 (0,52-0,94)	0,018	0,56 (0,28-1,10)	0,093
	rs3214012	1,12 (0,75-1,66)	0,584	1,14 (0,81-1,59)	0,458	1,32 (0,98-1,80)	0,072	1,63 (0,84-3,16)	0,152
	rs1801376	1,09 (0,72-1,64)	0,689	1,11 (0,78-1,57)	0,572	1,36 (0,98-1,88)	0,063	1,70 (0,87-3,35)	0,123
	rs2305653	1,07 (0,73-1,59)	0,727	1,13 (0,81-1,56)	0,481	1,46 (1,07-1,99)	0,016	2,30 (1,12-4,73)	0,023
	rs4924437	1,38 (0,90-2,12)	0,135	1,17 (0,83-1,66)	0,367	1,10 (0,81-1,51)	0,537	1,21 (0,60-2,44)	0,592
<i>c-MYC</i>	rs2070583	0,61 (0,20-1,84)	0,377	1,83 (0,63-5,33)	0,265	1,32 (0,53-3,30)	0,555	1,18 (0,15-9,32)	0,874
<i>EP300</i>	rs20552	0,83 (0,55-1,23)	0,343	1,16 (0,82-1,63)	0,407	1,25 (0,91-1,71)	0,179	1,82 (0,88-3,77)	0,106
<i>MDM2</i>	rs1144943	1,05 (0,72-1,55)	0,795	1,25 (0,90-1,73)	0,180	1,14 (0,85-1,52)	0,387	0,64 (0,31-1,31)	0,220
	rs937282	0,99 (0,69-1,43)	0,954	1,23 (0,90-1,67)	0,189	1,10 (0,83-1,45)	0,507	0,71 (0,37-1,36)	0,305
	rs937283	1,12 (0,75-1,66)	0,594	1,26 (0,90-1,77)	0,176	1,12 (0,83-1,51)	0,456	0,59 (0,27-1,27)	0,178
	rs1470383	1,03 (0,63-1,70)	0,897	0,92 (0,60-1,40)	0,690	1,03 (0,70-1,51)	0,888	1,42 (0,64-3,16)	0,390
	rs1201644	0,97 (0,66-1,43)	0,881	1,17 (0,85-1,61)	0,328	1,06 (0,80-1,40)	0,693	0,78 (0,41-1,50)	0,459
	rs3730590	1,02 (0,69-1,52)	0,905	1,22 (0,87-1,70)	0,253	1,04 (0,77-1,40)	0,819	0,72 (0,35-1,48)	0,367
	rs1695144	0,86 (0,22-3,33)	0,829	1,35 (0,39-4,69)	0,641	1,14 (0,44-2,95)	0,792	1,28 (0,16-10,16)	0,813
	rs3730646	1,04 (0,70-1,54)	0,844	1,23 (0,89-1,71)	0,215	1,10 (0,82-1,47)	0,543	0,72 (0,35-1,48)	0,371
	rs3730654	-	-	-	-	-	-	-	-
<i>RB1</i>	rs9568029	1,95 (1,19-3,21)	0,008	1,32 (0,91-1,92)	0,141	1,18 (0,84-1,65)	0,348	0,98 (0,46-2,08)	0,960
	rs1573601	2,00 (1,19-3,40)	0,010	1,30 (0,88-1,91)	0,193	1,38 (0,99-1,93)	0,061	0,66 (0,27-1,62)	0,368
	rs2804094	1,13 (0,52-2,44)	0,764	0,97 (0,52-1,79)	0,909	0,67 (0,35-1,26)	0,210	1,22 (0,36-4,20)	0,753
	rs4151441	1,51 (0,32-7,13)	0,605	0,65 (0,21-1,98)	0,448	0,32 (0,07-1,43)	0,134	-	-
	rs520342	1,79 (1,07-2,98)	0,027	1,29 (0,87-1,90)	0,206	1,39 (1,00-1,94)	0,053	0,68 (0,28-1,64)	0,389
	rs4151474	1,49 (0,31-7,04)	0,617	0,65 (0,21-1,96)	0,440	0,32 (0,07-1,44)	0,136	-	-
	rs399413	1,76 (1,12-2,77)	0,015	1,18 (0,84-1,66)	0,344	1,15 (0,84-1,57)	0,383	0,66 (0,30-1,46)	0,305
	rs9568036	1,27 (0,85-1,91)	0,244	1,07 (0,76-1,50)	0,710	1,03 (0,75-1,40)	0,874	0,49 (0,22-1,11)	0,086
	rs1894255	1,88 (1,13-3,14)	0,015	1,20 (0,82-1,76)	0,341	1,33 (0,95-1,86)	0,094	0,52 (0,29-1,35)	0,176
	rs1951775	1,71 (1,09-2,66)	0,019	1,22 (0,86-1,71)	0,261	1,11 (0,81-1,52)	0,504	0,75 (0,35-1,61)	0,457
	rs198575	1,15 (0,52-2,55)	0,723	1,00 (0,53-1,87)	0,992	0,60 (0,31-1,15)	0,125	1,24 (0,35-4,40)	0,736
	rs198577	1,22 (0,53-2,79)	0,638	1,12 (0,58-2,17)	0,734	0,53 (0,26-1,08)	0,080	1,34 (0,38-4,76)	0,650
	rs198580	0,99 (0,40-2,43)	0,983	1,14 (0,54-2,40)	0,726	0,70 (0,33-1,49)	0,359	1,84 (0,52-6,60)	0,348
	rs198590	1,75 (1,07-2,88)	0,027	1,19 (0,82-1,74)	0,357	1,31 (0,94-1,83)	0,107	0,62 (0,25-1,51)	0,288
	rs367438	1,19 (0,52-2,73)	0,680	1,07 (0,55-2,08)	0,834	0,62 (0,31-1,23)	0,168	1,39 (0,39-4,94)	0,609
<i>RBL2/p130</i>	rs10748	0,84 (0,57-1,23)	0,370	0,89 (0,65-1,22)	0,464	1,02 (0,77-1,36)	0,870	0,54 (0,26-1,10)	0,088
<i>STK6</i>	rs1468056	1,06 (0,74-1,52)	0,749	0,85 (0,62-1,15)	0,284	1,08 (0,82-1,44)	0,584	0,80 (0,41-1,57)	0,512
	rs2273535	0,80 (0,49-1,30)	0,360	0,96 (0,64-1,43)	0,828	1,05 (0,73-1,53)	0,780	0,98 (0,41-2,37)	0,963
	rs1047972	0,98 (0,60-1,62)	0,942	0,99 (0,65-1,49)	0,947	1,53 (1,06-2,23)	0,025	0,62 (0,22-1,77)	0,368
	rs6014711	0,85 (0,51-1,41)	0,526	0,70 (0,46-1,06)	0,094	1,42 (0,97-2,10)	0,074	0,68 (0,24-1,96)	0,474
	rs911160	1,14 (0,74-1,77)	0,548	1,20 (0,84-1,70)	0,320	0,98 (0,71-1,36)	0,909	0,90 (0,42-1,93)	0,786
	rs1044377	1,02 (0,69-1,51)	0,916	1,15 (0,83-1,58)	0,401	1,21 (0,91-1,62)	0,192	0,74 (0,36-1,54)	0,422
<i>VDR</i>	rs4516035	0,96 (0,66-1,41)	0,852	0,92 (0,67-1,26)	0,601	0,93 (0,70-1,23)	0,611	0,53 (0,25-1,11)	0,093
	rs10735810	1,24 (0,84-1,84)	0,280	1,12 (0,80-1,55)	0,513	0,95 (0,71-1,27)	0,721	1,00 (0,51-1,96)	0,991
	rs731236	1,31 (0,90-1,90)	0,157	0,91 (0,67-1,24)	0,555	0,94 (0,71-1,25)	0,686	1,20 (0,63-2,27)	0,581
	rs739837	1,02 (0,71-1,46)	0,912	1,12 (0,83-1,51)	0,480	1,09 (0,82-1,45)	0,544	1,11 (0,57-2,17)	0,750

Análisis estadístico del efecto aditivo del genotipo (en su alelo menor) para los genes no *E2F* relacionados con el control del ciclo celular, entre los diferentes grupos de muestras segregadas en función de las características tumorales: tumor ER positivo *versus* tumor ER negativo, tumor PR positivo *versus* tumor PR negativo, tumor metastatizado a ganglio linfático axilar *versus* tumor no metastatizado a ganglio linfático axilar, y tumor metastatizado a otros tejidos distales *versus* tumor no metastatizado a otros tejidos distales. Los resultados estadísticamente significativos se indican en negrita rojo. Los resultados marginales se indican en negrita negro.

V. DISCUSIÓN

V.1. EL GEN *FANCD2* MUESTRA UNA ASOCIACIÓN CON EL CÁNCER DE MAMA ESPORÁDICO

V.1.1. Un cambio sinónimo en *FANCD2* presenta asociación con el cáncer de mama esporádico

El gen *FANCD2* se había relacionado con el cáncer de mama en varias ocasiones, anteriormente al trabajo presentado en esta tesis doctoral. Así, se describió el incremento de expresión de *FANCD2* en tejidos con una alta tasa de proliferación, como ocurre en el tejido mamario (Holzel *et al.*, 2003), y estudios realizados en ratones mostraron cómo alteraciones de *Fancd2* murino en línea germinal producían cáncer de mama y ovario (Houghtaling *et al.*, 2005; Houghtaling *et al.*, 2003). Sin embargo, ningún grupo había propuesto a *FANCD2* como un gen de baja susceptibilidad para cáncer de mama esporádico hasta la fecha de publicación de nuestro trabajo y sólo un trabajo posterior establece la implicación de *FANCD2* en la patología. En éste, van der Groep y colaboradores describen, gracias a experimentos de inmunohistoquímica, que un 10-20% de los tumores de cáncer de mama esporádico y hereditario con mutación en *BRCA1* presentaban ausencia de actividad de la proteína *FANCD2*. Ello podría indicar que cambios presentes en *FANCD2* o la inactivación epigenética del gen tienen una gran importancia en la carcinogénesis del cáncer de mama (van der Groep *et al.*, 2008). En el presente trabajo se ha investigado la asociación entre las dos variantes inicialmente seleccionadas en el gen *FANCD2*, rs2272125 y rs722509, y el cáncer de mama esporádico. Se identificó, así, un incremento en el riesgo a padecer cáncer de mama esporádico asociado al alelo menor “C” del SNP sinónimo rs2272125 (p.Leu1366Leu) (OR por alelo=1,36, p=0,006). Con el fin de evidenciar la implicación de este cambio en la asociación a riesgo, se realizaron análisis con los haplotipos y diplotipos estimados para ambos SNPs genotipados en *FANCD2*, lo que nos permitió confirmar dicha asociación: los portadores de al menos una copia del haplotipo “TC”, portador del alelo menor “C” de rs2272125, presentaban un riesgo a cáncer de mama cuatro veces mayor que los no portadores (p=0,007 y p=0,006 para haplotipos y diplotipos portadores del alelo “C”, respectivamente). Inicialmente se tomó en consideración la posibilidad de que el SNP rs2272125 asociado con el cáncer de mama esporádico fuera el agente causal del riesgo, a pesar de ser un SNP sinónimo, pues son varias las formas en las que un cambio de

estas características puede intervenir directamente en la etiología de la enfermedad. Una de ellas es la implicación del cambio en motivos reguladores en *cis*, como son las regiones activadoras de “*splicing*” (ESE, *exon splicing enhancer*) o regiones silenciadoras de “*splicing*” (ESS, *exon splicing silencer*). En éstas, un SNP produce una secuencia diferente que no va a ser reconocida de igual modo por la maquinaria de “*splicing*” y que lleva a un peor o mejor rendimiento (Nielsen *et al.*, 2007), o incluso a la creación de un “*splicing*” alternativo que produce una proteína no funcional (Blencowe, 2000). Sin embargo, en nuestro caso se descartó la implicación de este cambio en un “*splicing*” aberrante del ARNm de *FANCD2* al no encontrarse dentro de una región de estas características tras un análisis *in silico*. Otra propuesta a la causalidad del SNP en *FANCD2* asociado a cáncer de mama ha sido que los cambios sinónimos pueden influir en la correcta producción de la cantidad de proteína aunque no lleven a un cambio en la cadena de aminoácidos (Chamary and Hurst, 2005; Kahali *et al.*, 2007; Schattner and Diekhans, 2006; Shabalina *et al.*, 2006). Sin embargo, la aplicación de esta base teórica en la determinación de la implicación de rs2272125 a través de ensayos funcionales no ha sido posible debido a la falta de material biológico disponible. Por ello, se consideró necesaria la búsqueda de otra u otras variantes que pudieran tener una implicación más directa en la patología con el fin de apoyar los resultados de asociación detectados.

V.1.2. Un SNP relacionado con el cambio asociado a cáncer de mama se encuentra en una región reguladora de la expresión génica de *FANCD2*

Se propuso que la asociación detectada podría tener su explicación en un cambio causal próximo a rs2272125 y en desequilibrio de ligamiento (DL) con éste. Para ello, se llevó a cabo un estudio en profundidad de las regiones de interés de *FANCD2* entre los portadores del haplotipo de riesgo “TC”, lo que permitió detectar la presencia de ocho variantes, algunas de ellas previamente identificadas y otras detectadas por primera vez. Se hipotetizó que cuatro de estas variantes podrían tener implicaciones funcionales en la expresión de *FANCD2* debido a su localización: región promotora, sitio de unión exón/intrón y 3’UTR. Además, cuatro de ellas se agruparon en una forma variante del haplotipo “TC”, denominada “H2V”. El análisis de la frecuencia de estas variantes definidas por “H2V” en personas no portadoras de “TC” mostró que no se asociaba a este cambio con el cáncer de mama ($p=0,200$), si bien una de las dos variantes

localizada en la posible región promotora de *FANCD2*, rs33917318, se encuentra en un motivo altamente conservado en la línea de mamíferos y presenta secuencias consenso de unión de factores de transcripción. Esto hizo especular si este cambio podría tener función en la expresión del gen. Así pues, si bien no se descarta la posibilidad de que este SNP en el promotor del gen esté relacionado de algún modo con el desarrollo del cáncer de mama, es posible que otra u otras variantes en el gen *FANCD2* o en regiones próximas al mismo sean las responsables del incremento de riesgo asociado a cáncer de mama detectado (Barroso *et al.*, 2006).

V.2. LA FAMILIA DE GENES *FANC*, LOS GENES RELACIONADOS CON ÉSTA Y SU ASOCIACIÓN CON EL CÁNCER DE MAMA ESPORÁDICO

Estudios previos han demostrado que varios genes involucrados en la reparación del daño en el ADN, diferentes de *BRCA1* y *BRCA2*, se relacionan con susceptibilidad a cáncer de mama, como son *ATM*, *CHEK2* y *P53*, entre otros (Consortium, 2004; de Jong *et al.*, 2002; Easton, 1994; Meijers-Heijboer *et al.*, 2002; Renwick *et al.*, 2006; Rodriguez-Lopez *et al.*, 2003; Smith *et al.*, 2003; Thorstenson *et al.*, 2003). Como se ha indicado en apartados anteriores, *BRCA2* ha sido identificado como *FANCD1* (Howlett *et al.*, 2002). Además, se han descrito ampliamente las interacciones que ocurren entre *BRCA1*, *BRCA2/FANCD1* y la familia de proteínas *FANC* (Folias *et al.*, 2002; Garcia and Benitez, 2008; Hussain *et al.*, 2004; Hussain *et al.*, 2003; Levrán *et al.*, 2005). Por tanto, se consideró plausible que alteraciones en los genes *FANC* estuvieran implicadas en un mayor riesgo a padecer cáncer de mama (Venkitaraman, 2002). Esta hipótesis también se sustenta en el hecho de que se han detectado alteraciones en los genes *FANC* en diferentes tipos de cáncer en la población general que no padece el síndrome de anemia de Fanconi, entre ellos el cáncer de mama (Alter *et al.*, 2003; Colleu-Durel *et al.*, 2001). Así, por ejemplo, se ha detectado la ausencia de actividad *FANCF* debido a la metilación aberrante del promotor del gen en cáncer de cérvix y cáncer de ovario (Narayan *et al.*, 2004; Wang *et al.*, 2006), y se han descrito mutaciones en *FANCC* y *FANCG* en cáncer de páncreas (van der Heijden *et al.*, 2003). A pesar de que la literatura indicaba el posible potencial de susceptibilidad a cáncer de mama y otros cánceres para la familia de genes *FANC*, pocos son los estudios que han investigado este tipo de asociaciones. Por tanto,

nuestro grupo consideró de interés la descripción de la asociación de variantes seleccionadas en los genes de la familia *FANC* y genes cuyas proteínas se relacionan con ésta, con el cáncer de mama esporádico. Se han obtenido varios resultados de asociación con el cáncer de mama esporádico para el grupo de genes indicado, fundamentalmente tras la estratificación de la muestra acorde a diversas características clínicas y tumorales. Sin embargo, dichos resultados requieren ser tratados con cautela puesto que todos muestran una significación estadística marginal, la mayoría de los cambios no presentan una implicación funcional en la proteína *a priori* y no se ha realizado un estudio funcional asociado o una replicación en una población independiente que los certifique como SNPs asociados a la patología.

V.2.1. Los miembros de la familia de genes *FANC* y los genes relacionados con *FANC* que presentan asociación directa con el cáncer de mama esporádico son los clásicos *BRCA1*, *BRCA2/FANCD1* y *ATM*

En *BRCA1* se asoció el cambio no sinónimo rs16941 (p.Gly1038Glu) con una tendencia a un mayor riesgo de padecer cáncer de mama esporádico (OR=1,17, p=0,084), mientras que el cambio intrónico rs799923 mostró asociación a protección frente a la patología (OR=0,82, p=0,048). Sin embargo, rs16941 se asoció de un modo significativo con la enfermedad entre las pacientes que presentaron tumores *in situ* (p=0,004), junto con otros dos cambios en el gen, rs8176265 (intrónico, p=0,010) y rs8176318 (3'UTR, p=0,006), que también se presentan de forma más frecuente entre las pacientes con tumores de tamaño menor a 2 cm (p=0,043 y p=0,013, respectivamente). En *BRCA2/FANCD1*, por su parte, se identificaron las variantes rs206146 (intrónica) y rs15869 (en 3'UTR) como cambios asociados a padecer cáncer de mama, con un riesgo estimado no superior a 1,25 (p=0,039 y p=0,033, respectivamente). Estos cambios, tras el análisis de estratificación de la muestra acorde a características clínicas y tumorales, no mostraron asociación a ningún grupo de forma específica. Son pocos los trabajos que han propuesto el estudio de variantes comunes en *BRCA1* y *BRCA2/FANCD1* en casos de cáncer de mama esporádico, dado que se ha centrado la atención en los casos de cáncer de mama hereditario (Palacios *et al.*, 2008) y ningún trabajo ha mostrado asociación a cáncer de mama para los SNPs arriba indicados. El estudio ya clásico de Freedman y colaboradores en 2004, determinó que, si bien las

variantes seleccionadas en *BRCA2/FANCD1* en su trabajo no ejercían un efecto evidente en el riesgo para la enfermedad, el haplotipo que formaban sí presentaba una asociación a un mayor riesgo de cáncer de mama esporádico (Freedman *et al.*, 2004). En un trabajo reciente desarrollado en población asiática, por su parte, se ha mostrado la asociación del haplotipo portador de una variante localizada en el promotor de *BRCA1* y protección a cáncer de mama esporádico entre mujeres mayores de 45 años, apoyado por un estudio funcional que pone en evidencia una mayor actividad del promotor del gen (Chan *et al.*, 2009). Sin embargo, el campo de estudio de *BRCA1* y *BRCA2/FANCD1* en el cáncer de mama esporádico sigue siendo limitado y requiere, como indican nuestros resultados, una mayor atención. En el gen *ATM* también se ha podido describir la asociación de un SNP intrónico, rs637064, con un mayor riesgo para el cáncer de mama esporádico. Éste muestra un riesgo estimado de 1,19 por alelo ($p=0,041$), con una asociación más marcada entre los pacientes con ausencia de metástasis a ganglios linfáticos mamarios ($p=0,008$). En la literatura se ha relacionado frecuentemente a *ATM* con el cáncer de mama familiar al proponer que aquellas mutaciones en homocigosis responsables del síndrome de ataxia telangiectasia (AT) para el que se describió inicialmente el gen, si se presentan en heterocigosis podrían estar involucradas en el desarrollo de diversos tipos de cáncer (Ahmed and Rahman, 2006). En 2006, Renwick y colaboradores describieron cómo dos mutaciones en *ATM* que causan comúnmente AT, en heterocigosis están más frecuentemente presentes entre un grupo de pacientes con cáncer de mama familiar negativas para mutaciones en *BRCA1* y *BRCA2/FANCD1*, que entre controles de población normal (Renwick *et al.*, 2006). Así como ocurre con este trabajo, otros han sido los que han relacionado a *ATM* con el riesgo a padecer cáncer de mama familiar (Chenevix-Trench *et al.*, 2002; Koren *et al.*, 2006; Pylkas *et al.*, 2007), pero pocos lo han propuesto como gen de susceptibilidad a riesgo para cáncer de mama esporádico (Bogdanova *et al.*, 2008; Stredrick *et al.*, 2006; Tamimi *et al.*, 2004) y ninguno ha analizado todas las características clínicas y tumorales que se han tratado en la presente tesis doctoral (Pooley *et al.*, 2008). Incluso existen resultados contradictorios en los que un cambio no sinónimo en *ATM* no sólo no se asocia con riesgo para cáncer de mama esporádico, sino que tiene un efecto protector (Schrauder *et al.*, 2008), o bien hay trabajos que no detectan asociación alguna con la enfermedad (Ye *et al.*, 2007). A

pesar de los resultados obtenidos para estos genes expuestos en la presente tesis, existen varios trabajos realizados en amplias poblaciones caucásicas que muestran resultados contradictorios a los nuestros. Baynes y colaboradores concluyeron que los *tag*-SNPs seleccionados en los genes *BRCA1*, *BRCA2/FANCD1*, *ATM*, *CHEK2* y *TP53*, para los que se han descrito mutaciones en cáncer de mama familiar, no se asocian en modo alguno con el cáncer de mama esporádico (Baynes *et al.*, 2007). Pharoah y colaboradores, por su parte, muestra la misma evidencia de falta de asociación con cáncer de mama esporádico para los genes *BRCA1*, *BRCA2/FANCD1* y *ATM*, analizados dentro de un estudio de 120 genes candidatos (Pharoah *et al.*, 2007). Asimismo, este último trabajo encuentra su continuación en el desarrollado por Mavaddat y colaboradores, donde se analizan las características clínicas y tumorales estudiadas también en nuestro trabajo en una amplia serie de casos y controles de cáncer de mama esporádico en relación a los genotipos obtenidos para los mismos 120 genes candidatos. En este trabajo se observa de nuevo que no hay evidencia de ninguna asociación de asociación para los tres genes que nos ocupan en este apartado (Mavaddat *et al.*, 2009), si bien estudios futuros adicionales que se realicen sobre los mismos SNPs estudiados en la presente tesis podrían aportar una mayor información.

V.2.2. Los miembros de la familia de genes *FANC* y los genes relacionados con *FANC* diferentes de *BRCA1*, *BRCA2/FANCD1* y *ATM* presentan asociación con el cáncer de mama esporádico tras la estratificación de la muestra

En el caso de los genes de la familia *FANC* y genes relacionados que no son *BRCA1*, *BRCA2/FANCD1* y *ATM*, se han encontrado asociaciones a la enfermedad únicamente tras la estratificación de la colección de muestras, lo que ha permitido homogeneizar las características tumorales, reducir la heterogeneidad de la muestra y, por tanto, definir mejor los grupos asociados al cáncer de mama (Garcia-Closas *et al.*, 2008; Palacios *et al.*, 2008; Sorlie *et al.*, 2003). Así por ejemplo, la mayoría de los SNPs seleccionados para *FANCA*, entre los que se encuentran tanto cambios intrónicos como codificantes no sinónimos, se asocian con el grupo de muestras que presentan una mejor supervivencia a la enfermedad, aunque un SNP intrónico en este gen mostró un comportamiento contrario a esta asociación, lo cual resta confianza a los resultados. Asimismo, un cambio en 3' corriente abajo de *FANCC*, que podría estar localizado en

regiones reguladoras de la expresión génica, fue asociado con un peor pronóstico para el cáncer de mama. Es necesario mencionar que, debido al reducido número de pacientes fallecidas por la enfermedad en nuestra colección, todos los resultados derivados de estos análisis hay que tomarlos con la cautela que ello implica. Existe únicamente un trabajo que evalúa a uno de los miembros de la familia de genes *FANC*, a *FANCD2*, en función de un valor pronóstico. En este trabajo identificaron un peor pronóstico en tumores invasivos con una alta proliferación celular y, por tanto, una alta dosis de expresión de *FANCD2* (van der Groep *et al.*, 2008). Otras han sido las asociaciones detectadas para el resto de características tumorales interrogadas para cambios presentes en seis miembros diferentes de la familia *FANC* (*FANCA*, *FANCC*, *FANCD2*, *FANCI*, *FANCI/BRIP1* y *FANCN/PALB2*) y en el gen *OGG1*, que codifica para la 8-oxo-guanina glicosilasa, una enzima implicada en la reparación de daños oxidativos del ADN como la generación de 8-oxo-guaninas, a través del sistema BER (*base-excision repair*) (Wood *et al.*, 2001). Se han descrito asociaciones a cáncer de mama esporádico al seleccionar pacientes con tumores *in situ*, con tumores de tamaño variable, positivos para PR, o que presentan metástasis a ganglios linfáticos axilares o metástasis a tejidos distales de la mama. Debido a que este es el primer trabajo que estudia la implicación de todos los genes de la familia *FANC* y de otros genes relacionados con ésta, con el cáncer de mama esporádico, no hay ningún otro estudio que apoye o desmienta las asociaciones a la clínica que aquí se presentan. Sin embargo, acorde a los resultados obtenidos, la profusión de asociaciones a propiedades clínicas contrapuestas para diversos SNPs en los genes indicados, apunta a que pueden existir cambios que potencien la actividad supresora de tumor de estas proteínas y que se verán relacionados con una clínica favorable, mientras que aquellos que interfieren de algún modo en la secuencia de la proteína o en la correcta expresión del gen, intervendrán en las situaciones contrarias. Sí existen algunos trabajos que han enfocado sus esfuerzos en determinar la implicación de *FANCI/BRIP1* y *FANCN/PALB2* en el cáncer de mama esporádico debido a su clara relación con el cáncer de mama familiar, aunque ninguno de estos trabajos ha analizado las características clínicas y tumorales señaladas (Chen *et al.*, 2008; De Nicolo *et al.*, 2008; Garcia *et al.*, 2008; Garcia *et al.*, 2009; Seal *et al.*, 2006; Tischkowitz *et al.*, 2007). Mavaddat y colaboradores, por su parte, no observaron

ningún resultado significativo para los genes *FANCI/BRIP1* y *ATR* que estudiaron (Mavaddat *et al.*, 2009), mientras Pooley y colaboradores, pudieron observar la asociación del SNP rs2191249 en *FANCI/BRIP1* con supervivencia a la enfermedad, si bien este cambio no ha sido incluido en nuestro grupo de SNPs (Pooley *et al.*, 2008). Así pues, de nuevo los resultados expuestos en este apartado requieren la confirmación a través de la replicación en una población independiente para la que se tengan datos clínicos y tumorales relevantes para el cáncer de mama.

V.3. EL GEN *E2F1* MUESTRA UNA ASOCIACIÓN CON EL CÁNCER DE MAMA ESPORÁDICO Y TIENE EFECTO EN LA EXPRESIÓN DEL GEN

La proteína E2F1 muestra un efecto bimodal, pues activa la apoptosis tras la señal de daño en el ADN, pero también actúa como un factor de transcripción para genes de proliferación celular (Polager *et al.*, 2002; Powers *et al.*, 2004). Este hecho ha evidenciado el importante papel que E2F1 puede tener en el cáncer, aunque todavía no esté claro si E2F1 promueve o protege frente a la carcinogénesis. Este efecto dual, que se considera dependiente de las rutas en las que intervienen pRB y p53 (Hallstrom *et al.*, 2008; Zacharatos *et al.*, 2004), también se ha visto reflejado en el comportamiento de E2F1 en diversos tipos de cánceres. Así, en el cáncer de vejiga y de mama, hay una correlación entre la sobre-expresión del gen *E2F1* y el índice de proliferación del tejido, mientras que en los cánceres de colon y próstata se observa el efecto contrario. Además, se ha visto que *BRCA1* tiene elementos diana para la actuación de E2F1 como factor de transcripción de su expresión génica (Wang *et al.*, 2000a). Acorde a estos datos, *E2F1* fue considerado un candidato apropiado para gen de baja penetrancia de susceptibilidad a cáncer de mama esporádico, entre otros miembros de la familia *E2F*.

V.3.1. Un cambio en 3'UTR de *E2F1* presenta asociación con el cáncer de mama esporádico entre pacientes mayores de 50 años

Nuestros análisis en el estudio caso-control realizado mostraron una asociación marginal del SNP rs3213180, localizado en 3'UTR de *E2F1* cuando se consideró la frecuencia alélica, genotípica, de haplotipos y de diplotipos. Sin embargo, esta asociación alcanzó la significación estadística únicamente cuando se consideraron las muestras

diagnosticadas para el cáncer de mama a una edad superior de 50 años de forma exclusiva, al realizar un estudio de estratificación de la muestra acorde a la edad de diagnóstico (OR=1,54, 95% I.C. 1,05-2,27, $p=0,027$). Este cambio fue inicialmente seleccionado debido a su apropiada frecuencia según los datos de la población HapMap CEU (12,6%), y por su localización en 3'UTR del gen y, por tanto, su posible consideración de SNP regulador o SNPr. Ambos factores son de gran importancia para el estudio de variantes de bajo riesgo en enfermedades complejas como es el cáncer (Aouacheria *et al.*, 2007; Chorley *et al.*, 2008; McCauley *et al.*, 2007; Prokunina and Alarcon-Riquelme, 2004). Sin embargo, en nuestro estudio, la frecuencia del alelo menor para rs3213180 cayó hasta un 5%, limitando, por tanto, el poder de obtener un resultado estadísticamente significativo en nuestro grupo de muestras.

V.3.2. El SNP rs3213180 interviene en la expresión génica de *E2F1*

Al tiempo de obtener los resultados de genotipado, surgió información acerca de la regulación de la expresión génica de *E2F1* a través del sistema de microRNAs. Así, se describió la actuación de los microRNAs miR-17-5p y miR-20a en la regulación de la trasducción del ARNm de *E2F1* a proteína, donde se encuentran dos sitios de unión, sitio 1 (S1) y sitio 2 (S2) en el extremo 3' no trasladado del ARNm (3'UTR, *3'untranslated region*) (O'Donnell *et al.*, 2005). Los microRNAs (miRNA) son ARNs pequeños (≈ 22 nucleótidos) no codificantes cuya misión es regular la traducción de sus ARNm diana mediante el reconocimiento y anillamiento específico de seis nucleótidos en la región 5' del miRNA, denominada "secuencia semilla" (Bartel, 2004; Lewis *et al.*, 2005). El mecanismo de bloqueo específico de la traducción requiere de la inclusión del miRNA dentro del complejo de silenciamiento denominado RISC (*RNA-induced silencing complex*), que está constituido por proteínas Argonautas, entre otras (Peters and Meister, 2007). El complejo formado es capaz de unirse a la secuencia complementaria del ARNm diana y promover su degradación o, más comúnmente en mamíferos, la represión transitoria de la traducción del ARNm a proteína en un momento y tejido específicos (Du and Zamore, 2005). La transcripción de los miRNAs miR-17-5p y miR-20a es activada por c-Myc, otro factor de transcripción que tiene a su cargo un efecto dual sobre la expresión de *E2F1*: estimula su transcripción, pero controla negativamente su traducción a través de la expresión de estos miRNAs (O'Donnell *et al.*, 2005). El cambio

V. DISCUSIÓN

V.3. El gen *E2F1* muestra una asociación con el cáncer de mama esporádico y tiene efecto en la expresión del gen

rs3213180 se ha visto que está próximo a S2, lo cual nos llevó a postular que rs3213180 podría afectar a la expresión de *E2F1* impidiendo la correcta unión de miR-17-5p y miR-20a a sus sitios específicos de unión al cambiar la estructura secundaria del ARNm. Se realizaron diversos experimentos utilizando el sistema de expresión *in vitro* de luciferasa, diseñando construcciones que portaban a 3'UTR de *E2F1* con S2 intacto o mutado, con el fin de testar la hipótesis anterior. Nuestros resultados mostraron que siempre que el alelo menor de rs3213180, "G", está presente en la forma completa de 3'UTR o en un fragmento del mismo incluyendo el SNP y S2, se observa un incremento de expresión al compararlo con las mismas construcciones portadoras del alelo normal "C" ($p < 0,001$). Esto sugiere algún tipo de efecto del SNP en la función de 3'UTR en la regulación de la expresión de *E2F1*. Esta idea encontró apoyo tras detectar que, si bien existen diferencias importantes en la expresión luciferasa para cada construcción en función del alelo de rs3213180, éstas no aparecen en la expresión del ARNm ($p = 0,835$), lo que apunta a un efecto del SNP como modulador de la expresión de *E2F1* a nivel de la traducción del ARNm. Asimismo, estudios de predicción del plegamiento del ARNm en ambas construcciones, 3'UTR completo y un fragmento del mismo, mostró que la estructura del ARNm cambia debido a la presencia del alelo "G". Sin embargo, esta hipótesis no ha podido ser confirmada de una manera evidente, puesto que en aquellas construcciones donde la función de S2 fue bloqueada tras la mutación de tres bases de la "secuencia semilla" de unión de los miRNAs, todavía se observaba un incremento en la expresión debido al alelo "G". Con el fin de darle más peso a nuestra hipótesis, se analizó si las muestras de tejido mamario normal de pacientes genotipadas para *E2F1* mostraban una expresión diferencial de la proteína E2F1 en función del alelo de rs3213180. Nuestros resultados mostraron una tendencia a un mayor número de núcleos E2F1 positivos entre las muestras homocigotas para "G" en rs3213180 ($p = 0,294$), mientras que la cuantificación de ARNm de estas mismas muestras mostró una cantidad igual de expresión del ARNm en los tres grupos de genotipos ($p = 0,913$). Aunque E2F1 es utilizado como marcador de proliferación en las células de cáncer de mama (Zhang *et al.*, 2000), estos resultados hay que tomarlos con cautela debido al bajo número de muestras que pudieron ser analizadas. Existe un estudio que apoya nuestra hipótesis y que describe el efecto de un SNP cercano a un sitio de unión de miRNAs en 3'UTR de un gen. Este

cambio se ha visto que interfiere claramente en la función de los miRNAs, y se produce la sobre-expresión de la dihidrofolato reductasa y la resistencia a metotrexato (Kertesz *et al.*, 2007; Mishra *et al.*, 2008; Mishra *et al.*, 2007). Este trabajo sugiere que el acceso al sitio de unión del miRNA por parte del complejo miRNA-RISC está bloqueado por la presencia del SNP. Acorde a esta idea, se podría considerar que rs3213180 ejerce un efecto similar en la expresión de *E2F1* a través del sistema de miRNAs. Sin embargo, la interferencia en la unión de factores de transcripción en 3'UTR de *E2F1* podría, asimismo, explicar el incremento en la expresión de *E2F1* y no debería ser descartada.

V.4. LA FAMILIA DE GENES *E2F*, OTROS GENES RELACIONADOS CON EL CONTROL DEL CICLO CELULAR Y SU ASOCIACIÓN CON EL CÁNCER DE MAMA ESPORÁDICO

V.4.1. El gen *VDR* porta dos SNPs que presentan asociación con el cáncer de mama esporádico

En el estudio de asociación de genes implicados en el control del ciclo celular diferentes de la familia de genes *E2F* y el cáncer de mama esporádico, se observaron diferencias en la MAF entre casos y controles para dos cambios del gen *VDR*. Dicho gen codifica el receptor de vitamina D₃, cuyo metabolito detectado, 1,25D (1 α ,25-dihydroxivitamin D₃ o calcitriol, la forma activa de la vitamina D₃) (Trang *et al.*, 1998), modula la expresión de genes específicos de proliferación celular, apoptosis y diferenciación de una forma tejido-específica y protege frente al cáncer de mama (Deeb *et al.*, 2007; Jurutka *et al.*, 2001). El cambio sinónimo rs731236 (p.Ile352Ile, *TaqI*) de *VDR*, presentó un OR estimado por alelo menor de 0,84 (95% I.C. 0,71-0,99, p=0,034), y se mantuvo en el análisis multivariante para los factores de confusión aplicados en esta tesis doctoral. El SNP rs731236 está en DL con otros polimorfismos en el extremo 3' del gen en poblaciones caucásicas como es la población española. Estudios funcionales de estos polimorfismos han evaluado la implicación putativa de la regulación de la transcripción, procesamiento del ARNm y traducción de *VDR*, sin resultados concluyentes (Durrin *et al.*, 1999). El cambio no sinónimo genotipado en *VDR*, rs2228570 (p.Met1Thr, *FokI*, rs10735810), por su parte, presentó una asociación no significativa con la enfermedad (OR=1,17, 95% I.C. 0,98-1,40, p=0,081), aunque llegó a la significación estadística tras la corrección por los factores de confusión (p=0,036). La diferencia entre

V. DISCUSIÓN

V.4. La familia de genes *E2F*, otros genes relacionados con el control del ciclo celular y su asociación con el cáncer de mama esporádico

ambos análisis podría ser debida a la estrecha relación entre la función de VDR y la etiología hormonal del cáncer de mama (Gilad *et al.*, 2005; Gilad and Schwartz, 2007). Estudios funcionales realizados en este SNP han descrito que el alelo menor “A” de rs2228570, parece estar asociado al uso de un codón de inicio alternativo al utilizado comúnmente, que produce una proteína más larga con un menor potencial de recepción de 1,25D (Arai *et al.*, 1997). Estudios previos realizados en VDR no han encontrado evidencia de asociación entre los dos SNPs mencionados arriba y el cáncer de mama (Abbas *et al.*, 2008; Bretherton-Watt *et al.*, 2001; Buyru *et al.*, 2003; Curran *et al.*, 1999; Hou *et al.*, 2002; John *et al.*, 2007; McCullough *et al.*, 2007). Sin embargo, estos estudios han mostrado un limitado poder estadístico de detección de la asociación moderada identificada en el presente trabajo, con resultados contradictorios para rs731236 (Abbas *et al.*, 2008; Curran *et al.*, 1999), y con asociación con riesgo para cáncer de mama en poblaciones caucásicas en el caso de rs2228570 (Chen *et al.*, 2005; Gapska *et al.*, 2008). Asimismo, estudios epidemiológicos han mostrado que existe asociación entre bajos niveles de 1,25D circulantes y el incremento del cáncer de mama, colorrectal y de próstata, mientras que se ha descrito la posible asociación entre variantes polimórficas en VDR y el incremento de cáncer primario o metastático en mama, colon y próstata (Bikle *et al.*, 2005; Cui and Rohan, 2006; Deeb *et al.*, 2007; Guy *et al.*, 2004), aunque su significado funcional sigue siendo desconocido. Esto puede explicarse debido a que el cáncer de mama podría estar, en parte, influenciado por efectos en la variación de la toma de vitamina D a través de la dieta y la exposición solar, en conjunción con los polimorfismos de VDR. A fin de encontrar implicaciones clínicas en los hallazgos de genotipado, se realizó el estudio de asociación entre los SNPs de VDR y varias características clínicas y tumorales, aunque no se llegó a ningún resultado concluyente (Barroso *et al.*, 2008).

V.4.2. Los miembros de la familia de genes *E2F* y otros genes relacionados con el control del ciclo celular presentan asociación con el cáncer de mama esporádico tras la estratificación de la muestra

Tras los resultados de carácter marginal en la asociación alélica a riesgo para cáncer de mama esporádico obtenidos para dos cambios que se encuentran en regiones putativas de regulación de la expresión génica en *E2F3*, rs4712494 (OR por alelo=1,30,

p=0,044), en el intrón 1, y rs1042317 (OR por alelo=1,18, p=0,054), en 3'UTR, se procedió a la estratificación de la muestra para éste y los demás genes de este grupo, en la búsqueda de una mejor definición de la asociación a cáncer de mama esporádico. Se pudo comprobar que rs4712494 se asocia a la enfermedad entre aquellas pacientes con un tumor de alto grado (p=0,001), lo cual puede indicar el potencial carcinogénico de este SNP localizado en el intrón 1 del gen, una región que, en numerosas ocasiones, interviene en la expresión génica junto con el promotor del gen. El SNP rs2328489, de localización intrónica en *E2F3*, muestra por su parte una asociación a la patología atendiendo a las muestras de menor grado tumoral (p=0,040). No se detectó ninguna otra asociación tras la estratificación de la muestra para ningún miembro de la familia *E2F* y, dado que no existen trabajos que apoyen las asociaciones detectadas, de nuevo la replicación en una población independiente se perfila como la mejor alternativa para la confirmación de las mismas. Asimismo, se han detectado asociaciones a cáncer de mama esporádico para SNPs localizados en los genes *BUB1B*, *STK6* y *RB1*, al considerar los grupos de muestras para las características clínicas de peor supervivencia a la enfermedad, pacientes diagnosticadas con tumores *in situ*, tumores de tamaño variable, tumores que presentan una expresión de ER positiva y para situaciones de presencia o ausencia de metástasis a ganglios linfáticos. La expresión génica de *BUB1B*, junto con la de *STK6*, está regulada por la proteína BRCA1, que ejerce su efecto en la progresión mitótica (Bae *et al.*, 2005) y, además, se ha destacado como un marcador que reconoce hasta un 80% del tejido tumoral mamario en parafina (Yuan *et al.*, 2006). De este modo, testar su implicación en la susceptibilidad a cáncer de mama esporádico se presentó interesante, aunque pocos trabajos han descrito este hecho (Lo *et al.*, 2007) y de nuevo, ninguno responde a un análisis de características clínicas y tumorales como el desarrollado en la presente tesis. *RB1*, por su parte, ha sido tradicionalmente considerado un gen involucrado en los eventos de carcinogénesis y su relación con el cáncer de mama ha sido descrita en varias ocasiones (Lesueur *et al.*, 2006), y trabajos con un mayor número de muestras que el nuestro han puesto de manifiesto la relación de este gen con el riesgo a padecer cáncer de mama (Mavaddat *et al.*, 2009), aunque ninguno de los dos SNPs asociados presentan dicha asociación en nuestro estudio. La aplicación de las características clínicas y tumorales en los estudios de asociación de

polimorfismos y cáncer de mama arriba indicados no es una práctica nueva, pues se pueden encontrar numerosos ejemplos en la literatura, la mayoría en estudios de genes candidatos como el que nos ocupa en la presente tesis doctoral (Bergman-Jungestrom *et al.*, 1999; Han *et al.*, 2004; Hong *et al.*, 2003; Rylander-Rudqvist *et al.*, 2003; Wedren *et al.*, 2003), si bien el que mejor caracteriza la patología hasta ahora es el realizado por Mavaddat y colaboradores (2009).

V.5. SNPs REGULADORES Y SU IMPLICACIÓN EN EL CÁNCER DE MAMA ESPORÁDICO

Muchos han sido los trabajos de asociación a cáncer de mama esporádico utilizando genes candidatos que se han publicado en los últimos años. Sin embargo, la mayoría de ellos se han enfocado tradicionalmente en SNPs que afectan de algún modo a la secuencia de la proteína, esto es, que reportan cambios no sinónimos, de pérdida de sentido o de cambio de sentido (Bergman-Jungestrom and Wingren, 2001; Egan *et al.*, 2003; Mitrinen *et al.*, 2001; Sazci *et al.*, 2004). Sin embargo, son pocos los que han sido replicados en estudios posteriores, y sólo estudios encuadrados dentro del Consorcio de Estudios de Asociación para Cáncer de Mama han llegado a resultados concluyentes para los genes *CASP8* y *TGFB1* aplicando el diseño de estudio de genes candidatos (Pharoah *et al.*, 2007). Estos estudios se presentan como los más representativos en este campo, para los cuales la significación estadística en un grupo muy amplio de muestras y la aplicación de estudios de asociación para SNPs localizados en todo el genoma avalan sus resultados. Éstos indican que los SNPs denominados SNPs reguladores (SNPr) pueden ejercer funciones de importancia en las enfermedades complejas como es el cáncer de mama esporádico. Así por ejemplo, el grupo de Easton y colaboradores detectó, a través de GWAS, la asociación de SNPs en cinco nuevas regiones con cáncer de mama esporádico que incluían regiones putativas reguladoras de los genes *FGFR2*, *TNRC9*, *LSP1*, *MAP3K1* y una región intergénica en 8q24, y tres de estos genes se relacionaron con el control del ciclo celular (Easton *et al.*, 2007). Ello ha abierto el campo al estudio de este tipo de genes que no se habían considerado hasta ahora para el cáncer de mama (Driver *et al.*, 2008). Estudios posteriores en amplias colecciones de muestras indicaron el mismo perfil de asociación de SNPr a cáncer de mama. Así por ejemplo, variantes localizadas en el intrón 2 de *FGFR2* se asociaron con cáncer de mama (Hunter *et al.*, 2007), cambios en el intrón 4 de *ESR1* se relacionaron con un mayor

riesgo para cáncer de mama (Dunning *et al.*, 2009), así como dos variantes en 16q12, en 5' de *TNRC9* (Stacey *et al.*, 2007). Las regiones reguladoras de los genes se consideran zonas de interés puesto que SNPs localizados en ellas, los SNPr, al tener un mayor MAF que las variantes codificantes (Chorley *et al.*, 2008; Prokunina and Alarcon-Riquelme, 2004) y al ejercer un efecto moderado en la expresión del gen, se postula que son los mayormente implicados en las patologías complejas. Entre las regiones reguladoras, las zonas promotoras, cambios sinónimos, intrones que regulan el "*splicing*" y 3'UTRs de los genes candidatos, se han considerado de interés como portadoras de SNPr, entre otras (Georges *et al.*, 2007; Saunders *et al.*, 2007; Zhu *et al.*, 2008). El hecho de que un SNP en una región promotora pueda estar asociado al cáncer de mama esporádico no es un concepto nuevo en la literatura, y muchos son los trabajos que han relacionado cambios en la región promotora de varios genes con una expresión alterada de las proteínas codificantes y con una asociación a cáncer de mama (Chan *et al.*, 2009; Rokavec *et al.*, 2007; Sun *et al.*, 2007). Asimismo, se han encontrado diversos SNPs en 3'UTR de genes candidatos de bajo riesgo para cáncer de mama que han presentado asociación a la patología (Brendle *et al.*, 2008; Sprague *et al.*, 2007), aunque en la mayoría de los casos la implicación funcional de estos SNPs en la expresión génica es desconocida. Como innovación en el estudio de SNPr, en la presente tesis doctoral hemos propuesto a los SNPr que se encuentran en 3'UTR no directamente localizados en las regiones de unión de miRNAs como cambios con una posible función en el correcto funcionamiento de este sistema. Esto encuentra apoyo en que los SNPs que se encuentran en los sitios de unión tienen una menor densidad y frecuencia que aquellos localizados fuera de los mismos al producir, en muchos casos, cambios deletéreos (Chen and Rajewsky, 2006; Yu *et al.*, 2007). Esto puede abrir un nuevo campo de búsqueda de SNPs causales de riesgo para enfermedades complejas, aparte de los cambios localizados en las regiones comúnmente relacionadas con la regulación génica arriba indicadas.

VI. CONCLUSIONES

A continuación se indican las conclusiones a las que se ha llegado con la consecución de la presente tesis doctoral:

1. Se ha descrito una asociación entre un SNP sinónimo en el exón 42 del gen *FANCD2*, rs2272125 (p.Leu1366Leu), y un riesgo incrementado de padecer cáncer de mama esporádico. La búsqueda de variantes en el mismo bloque de DL, putativas causales de la asociación a la patología detectada, mostró una mayor frecuencia del cambio rs33917318 en región promotora entre los casos con cáncer de mama, pero sin significación estadística.
2. El estudio completo de los genes de la familia *FANC* y genes cuyas proteínas se relacionan con esta familia de proteínas ha reportado la asociación de cambios intrónicos en los genes *BRCA1*, *BRCA2/FANCD1* y *ATM* con el cáncer de mama esporádico, mientras que la estratificación de la muestra relaciona a otros genes de la familia *FANC* con diversas características clínicas y tumorales, con un significado todavía por determinar.
3. El cambio rs3213180, localizado en 3'UTR de *E2F1*, presenta una tendencia a asociación con cáncer de mama esporádico que llega a la significación estadística entre las pacientes con una edad de diagnóstico mayor de 50 años. Dicho polimorfismo está implicado en la expresión de *E2F1* a nivel de su traducción a proteína, y la interrupción del sistema de microRNAs se perfila como el mecanismo de acción más probable.
4. Se ha detectado la asociación de dos SNPs en *VDR* con susceptibilidad a cáncer de mama esporádico. Uno de ellos, un cambio no sinónimo, rs2228570 (p.Met1Thr), se relaciona con un mayor riesgo para cáncer de mama y está respaldado por estudios funcionales previos en los que se describe que lleva a la producción de una proteína menos activa.
5. El estudio de otros genes que intervienen en el control del ciclo celular ha reportado la asociación de los genes *E2F3*, *BUB1B*, *STK6* y *RB1* con el cáncer de mama tras la estratificación de la muestra para diversas características clínicas y tumorales, cuyo significado es todavía desconocido.

6. Nuestros datos sugieren que:
 - a. La familia de genes *FANC* en su conjunto no parece conferir una mayor susceptibilidad al cáncer de mama esporádico.
 - b. Los genes de control del ciclo celular, que no habían sido considerados previamente en estudios de asociación a cáncer de mama esporádico, parecen despuntar como genes de relevancia en esta patología.
 - c. Los SNPs reguladores localizados en regiones reguladoras de la expresión génica como son las regiones 5'UTR, 3'UTR, reguladoras del "*splicing*" y codificantes con cambios sinónimos, pueden aportar un efecto leve en la expresión del gen e intervenir en el desarrollo de enfermedades de carácter complejo como es el cáncer de mama esporádico.

VII. BIBLIOGRAFÍA

Referencias bibliográficas presentadas en orden alfabético.

- Abbas S, Nieters A, Linseisen J, Slinger T, Kropp S, Mutschelknauss EJ *et al* (2008). Vitamin D receptor gene polymorphisms and haplotypes and postmenopausal breast cancer risk. *Breast Cancer Res* **10**: R31.
- Ahmed M, Rahman N (2006). ATM and breast cancer susceptibility. *Oncogene* **25**: 5906-11.
- Alter BP, Greene MH, Velazquez I, Rosenberg PS (2003). Cancer in Fanconi anemia. *Blood* **101**: 2072.
- Andreassen PR, D'Andrea AD, Taniguchi T (2004). ATR couples FANCD2 monoubiquitination to the DNA-damage response. *Genes Dev* **18**: 1958-63.
- Angele S, Romestaing P, Moullan N, Vuillaume M, Chapot B, Friesen M *et al* (2003). ATM haplotypes and cellular response to DNA damage: association with breast cancer risk and clinical radiosensitivity. *Cancer Res* **63**: 8717-25.
- Antoniou A, Pharoah PD, Narod S, Risch HA, Eyfjord JE, Hopper JL *et al* (2003). Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. *Am J Hum Genet* **72**: 1117-30.
- Antoniou AC, Easton DF (2003). Polygenic inheritance of breast cancer: Implications for design of association studies. *Genet Epidemiol* **25**: 190-202.
- Antoniou AC, Pharoah PD, McMullan G, Day NE, Ponder BA, Easton D (2001). Evidence for further breast cancer susceptibility genes in addition to BRCA1 and BRCA2 in a population-based study. *Genet Epidemiol* **21**: 1-18.
- Antoniou AC, Sinilnikova OM, Simard J, Leone M, Dumont M, Neuhausen SL *et al* (2007). RAD51 135G-->C modifies breast cancer risk among BRCA2 mutation carriers: results from a combined analysis of 19 studies. *Am J Hum Genet* **81**: 1186-200.
- Aouacheria A, Navratil V, Lopez-Perez R, Gutierrez NC, Churkin A, Barash D *et al* (2007). In silico whole-genome screening for cancer-related single-nucleotide polymorphisms located in human mRNA untranslated regions. *BMC Genomics* **8**: 2.
- Arai H, Miyamoto K, Taketani Y, Yamamoto H, Iemori Y, Morita K *et al* (1997). A vitamin D receptor gene polymorphism in the translation initiation codon: effect on protein activity and relation to bone mineral density in Japanese women. *J Bone Miner Res* **12**: 915-21.
- Attwooll C, Lazzerini Denchi E, Helin K (2004). The E2F family: specific functions and overlapping interests. *Embo J* **23**: 4709-16.
- Bae I, Rih JK, Kim HJ, Kang HJ, Haddad B, Kirilyuk A *et al* (2005). BRCA1 regulates gene expression for orderly mitotic progression. *Cell Cycle* **4**: 1641-66.
- Barrett JC, Fry B, Maller J, Daly MJ (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**: 263-5.
- Barroso E, Fernandez LP, Milne RL, Pita G, Sendagorta E, Floristan U *et al* (2008). Genetic analysis of the vitamin D receptor gene in two epithelial cancers: melanoma and breast cancer case-control studies. *BMC Cancer* **8**: 385.
- Barroso E, Milne RL, Fernandez LP, Zamora P, Arias JI, Benitez J *et al* (2006). FANCD2 associated with sporadic breast cancer risk. *Carcinogenesis* **27**: 1930-7.
- Bartel DP (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**: 281-97.

- Baynes C, Healey CS, Pooley KA, Scollen S, Luben RN, Thompson DJ *et al* (2007). Common variants in the ATM, BRCA1, BRCA2, CHEK2 and TP53 cancer susceptibility genes are unlikely to increase breast cancer risk. *Breast Cancer Res* **9**: R27.
- Bergman-Jungestrom M, Gentile M, Lundin AC, Wingren S (1999). Association between CYP17 gene polymorphism and risk of breast cancer in young women. *Int J Cancer* **84**: 350-3.
- Bergman-Jungestrom M, Wingren S (2001). Catechol-O-Methyltransferase (COMT) gene polymorphism and breast cancer risk in young women. *Br J Cancer* **85**: 859-62.
- Bikle DD, Oda Y, Xie Z (2005). Vitamin D and skin cancer: a problem in gene regulation. *J Steroid Biochem Mol Biol* **97**: 83-91.
- Blencowe BJ (2000). Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem Sci* **25**: 106-10.
- Bochar DA, Wang L, Beniya H, Kinev A, Xue Y, Lane WS *et al* (2000). BRCA1 is associated with a human SWI/SNF-related complex: linking chromatin remodeling to breast cancer. *Cell* **102**: 257-65.
- Bodvarsdottir SK, Hilmarsdottir H, Birgisdottir V, Steinarsdottir M, Jonasson JG, Eyfjord JE (2007). Aurora-A amplification associated with BRCA2 mutation in breast tumours. *Cancer Lett* **248**: 96-102.
- Bogdanova N, Cybulski C, Bermisheva M, Datsyuk I, Yamini P, Hillemanns P *et al* (2008). A nonsense mutation (E1978X) in the ATM gene is associated with breast cancer. *Breast Cancer Res Treat*.
- Bogliolo M, Lyakhovich A, Callen E, Castella M, Cappelli E, Ramirez MJ *et al* (2007). Histone H2AX and Fanconi anemia FANCD2 function in the same pathway to maintain chromosome stability. *Embo J* **26**: 1340-51.
- Bond GL, Hu W, Bond EE, Robins H, Lutzker SG, Arva NC *et al* (2004). A single nucleotide polymorphism in the MDM2 promoter attenuates the p53 tumor suppressor pathway and accelerates tumor formation in humans. *Cell* **119**: 591-602.
- Bonnen PE, Wang PJ, Kimmel M, Chakraborty R, Nelson DL (2002). Haplotype and linkage disequilibrium architecture for human cancer-associated genes. *Genome Res* **12**: 1846-53.
- Bradbury AR, Olopade OI (2007). Genetic susceptibility to breast cancer. *Rev Endocr Metab Disord* **8**: 255-67.
- Brendle A, Lei H, Brandt A, Johansson R, Enquist K, Henriksson R *et al* (2008). Polymorphisms in predicted microRNA-binding sites in integrin genes and breast cancer: ITGB4 as prognostic marker. *Carcinogenesis* **29**: 1394-9.
- Bretherton-Watt D, Given-Wilson R, Mansi JL, Thomas V, Carter N, Colston KW (2001). Vitamin D receptor gene polymorphisms are associated with breast cancer risk in a UK Caucasian population. *Br J Cancer* **85**: 171-5.
- Bridge WL, Vandenberg CJ, Franklin RJ, Hiom K (2005). The BRIP1 helicase functions independently of BRCA1 in the Fanconi anemia pathway for DNA crosslink repair. *Nat Genet* **37**: 953-7.
- Buyru N, Tezol A, Yosunkaya-Fenerci E, Dalay N (2003). Vitamin D receptor gene polymorphisms in breast cancer. *Exp Mol Med* **35**: 550-5.
- Campbell IG, Choong D, Chenevix-Trench G (2004). No germline mutations in the histone acetyltransferase gene EP300 in BRCA1 and BRCA2 negative families with breast cancer and gastric, pancreatic, or colorectal cancer. *Breast Cancer Res* **6**: R366-71.

- Cantor SB, Bell DW, Ganesan S, Kass EM, Drapkin R, Grossman S *et al* (2001). BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function. *Cell* **105**: 149-60.
- Carter RF (2001). BRCA1, BRCA2 and breast cancer: a concise clinical review. *Clin Invest Med* **24**: 147-57.
- Cavalieri E, Frenkel K, Liehr JG, Rogan E, Roy D (2000). Estrogens as endogenous genotoxic agents--DNA adducts and mutations. *J Natl Cancer Inst Monogr*: 75-93.
- Chamary JV, Hurst LD (2005). Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. *Genome Biol* **6**: R75.
- Chan KY, Liu W, Long JR, Yip SP, Chan SY, Shu XO *et al* (2009). Functional polymorphisms in the BRCA1 promoter influence transcription and are associated with decreased risk for breast cancer in Chinese women. *J Med Genet* **46**: 32-9.
- Chen K, Rajewsky N (2006). Natural selection on human microRNA binding sites inferred from SNP data. *Nat Genet* **38**: 1452-6.
- Chen P, Liang J, Wang Z, Zhou X, Chen L, Li M *et al* (2008). Association of common PALB2 polymorphisms with breast cancer risk: a case-control study. *Clin Cancer Res* **14**: 5931-7.
- Chen S, Iversen ES, Friebel T, Finkelstein D, Weber BL, Eisen A *et al* (2006). Characterization of BRCA1 and BRCA2 mutations in a large United States sample. *J Clin Oncol* **24**: 863-71.
- Chen WY, Bertone-Johnson ER, Hunter DJ, Willett WC, Hankinson SE (2005). Associations between polymorphisms in the vitamin D receptor and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* **14**: 2335-9.
- Chenevix-Trench G, Spurdle AB, Gatei M, Kelly H, Marsh A, Chen X *et al* (2002). Dominant negative ATM mutations in breast cancer families. *J Natl Cancer Inst* **94**: 205-15.
- Chorley BN, Wang X, Campbell MR, Pittman GS, Nouredine MA, Bell DA (2008). Discovery and verification of functional single nucleotide polymorphisms in regulatory genomic regions: current and developing technologies. *Mutat Res* **659**: 147-57.
- Chrzan P, Skokowski J, Karmolinski A, Pawelczyk T (2001). Amplification of c-myc gene and overexpression of c-Myc protein in breast cancer and adjacent non-neoplastic tissue. *Clin Biochem* **34**: 557-62.
- Classon M, Dyson N (2001). p107 and p130: versatile proteins with interesting pockets. *Exp Cell Res* **264**: 135-47.
- Classon M, Harlow E (2002). The retinoblastoma tumour suppressor in development and cancer. *Nat Rev Cancer* **2**: 910-7.
- Cliby WA, Roberts CJ, Cimprich KA, Stringer CM, Lamb JR, Schreiber SL *et al* (1998). Overexpression of a kinase-inactive ATR protein causes sensitivity to DNA-damaging agents and defects in cell cycle checkpoints. *Embo J* **17**: 159-69.
- Colleu-Durel S, Guitton N, Nourgalieva K, Leveque J, Danic B, Chenal C (2001). Genomic instability and breast cancer. *Oncol Rep* **8**: 1001-5.
- Comings DE, Gade-Andavolu R, Cone LA, Muhleman D, MacMurray JP (2003). A multigene test for the risk of sporadic breast carcinoma. *Cancer* **97**: 2160-70.
- Conde L, Vaquerizas JM, Dopazo H, Arbiza L, Reumers J, Rousseau F *et al* (2006). PupaSuite: finding functional single nucleotide polymorphisms for large-scale genotyping purposes. *Nucleic Acids Res* **34**: W621-5.

- Conde L, Vaquerizas JM, Santoyo J, Al-Shahrour F, Ruiz-Llorente S, Robledo M *et al* (2004). PupaSNP Finder: a web tool for finding SNPs with putative effect at transcriptional level. *Nucleic Acids Res* **32**: W242-8.
- Consortium BCA (2006). Commonly studied single-nucleotide polymorphisms and breast cancer: results from the Breast Cancer Association Consortium. *J Natl Cancer Inst* **98**: 1382-96.
- Consortium TCBC-C (2004). CHEK2*1100delC and susceptibility to breast cancer: a collaborative analysis involving 10,860 breast cancer cases and 9,065 controls from 10 studies. *Am J Hum Genet* **74**: 1175-82.
- Cotran RS, Kumar V, Collins T (2000). Robbins. Patología Estructural y Funcional. *McGraw-Hill*: 1137-1162.
- Cowey CL, Rathmell WK (2009). VHL gene mutations in renal cell carcinoma: role as a biomarker of disease outcome and drug efficacy. *Curr Oncol Rep* **11**: 94-101.
- Cox A, Dunning AM, Garcia-Closas M, Balasubramanian S, Reed MW, Pooley KA *et al* (2007). A common coding variant in CASP8 is associated with breast cancer risk. *Nat Genet* **39**: 352-8.
- Cox DG, Kraft P, Hankinson SE, Hunter DJ (2005). Haplotype analysis of common variants in the BRCA1 gene and risk of sporadic breast cancer. *Breast Cancer Res* **7**: R171-5.
- Cui Y, Rohan TE (2006). Vitamin D, calcium, and breast cancer risk: a review. *Cancer Epidemiol Biomarkers Prev* **15**: 1427-37.
- Curran JE, Vaughan T, Lea RA, Weinstein SR, Morrison NA, Griffiths LR (1999). Association of A vitamin D receptor polymorphism with sporadic breast cancer development. *Int J Cancer* **83**: 723-6.
- Cuzick J (2008). Assessing risk for breast cancer. *Breast Cancer Res* **10 Suppl 4**: S13.
- D'Andrea AD, Grompe M (2003). The Fanconi anaemia/BRCA pathway. *Nat Rev Cancer* **3**: 23-34.
- Danielian PS, Friesenhahn LB, Faust AM, West JC, Caron AM, Bronson RT *et al* (2008). E2f3a and E2f3b make overlapping but different contributions to total E2f3 activity. *Oncogene*.
- Davies AA, Masson JY, McIlwraith MJ, Stasiak AZ, Stasiak A, Venkitaraman AR *et al* (2001). Role of BRCA2 in control of the RAD51 recombination and DNA repair protein. *Mol Cell* **7**: 273-82.
- de Jong MM, Nolte IM, te Meerman GJ, van der Graaf WT, Oosterwijk JC, Kleibeuker JH *et al* (2002). Genes other than BRCA1 and BRCA2 involved in breast cancer susceptibility. *J Med Genet* **39**: 225-42.
- De Nicolo A, Tancredi M, Lombardi G, Flemma CC, Barbuti S, Di Cristofano C *et al* (2008). A novel breast cancer-associated BRIP1 (FANCI/BACH1) germ-line mutation impairs protein stability and function. *Clin Cancer Res* **14**: 4672-80.
- Deb SP (2003). Cell cycle regulatory functions of the human oncoprotein MDM2. *Mol Cancer Res* **1**: 1009-16.
- Deeb KK, Trump DL, Johnson CS (2007). Vitamin D signalling pathways in cancer: potential for anticancer therapeutics. *Nat Rev Cancer* **7**: 684-700.
- Digweed M, Rothe S, Demuth I, Scholz R, Schindler D, Stumm M *et al* (2002). Attenuation of the formation of DNA-repair foci containing RAD51 in Fanconi anaemia. *Carcinogenesis* **23**: 1121-6.
- Driver KE, Song H, Lesueur F, Ahmed S, Barbosa-Morais NL, Tyrer JP *et al* (2008). Association of single-nucleotide polymorphisms in the cell cycle genes with breast cancer in the British population. *Carcinogenesis* **29**: 333-41.

- Du T, Zamore PD (2005). microPrimer: the biogenesis and function of microRNA. *Development* **132**: 4645-52.
- Dunning AM, Healey CS, Baynes C, Maia AT, Scollen S, Vega A *et al* (2009). Association of ESR1 gene tagging SNPs with breast cancer risk. *Hum Mol Genet* **18**: 1131-9.
- Dunning AM, Healey CS, Pharoah PD, Teare MD, Ponder BA, Easton DF (1999). A systematic review of genetic polymorphisms and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* **8**: 843-54.
- DuPree EL, Mazumder S, Almasan A (2004). Genotoxic stress induces expression of E2F4, leading to its association with p130 in prostate carcinoma cells. *Cancer Res* **64**: 4390-3.
- Durrin LK, Haile RW, Ingles SA, Coetzee GA (1999). Vitamin D receptor 3'-untranslated region polymorphisms: lack of effect on mRNA stability. *Biochim Biophys Acta* **1453**: 311-20.
- Easton DF (1994). Cancer risks in A-T heterozygotes. *Int J Radiat Biol* **66**: S177-82.
- Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG *et al* (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**: 1087-93.
- Egan KM, Newcomb PA, Ambrosone CB, Trentham-Dietz A, Titus-Ernstoff L, Hampton JM *et al* (2004). STK15 polymorphism and breast cancer risk in a population-based study. *Carcinogenesis* **25**: 2149-53.
- Egan KM, Thompson PA, Titus-Ernstoff L, Moore JH, Ambrosone CB (2003). MnSOD polymorphism and breast cancer in a population-based case-control study. *Cancer Lett* **199**: 27-33.
- ElSharawy A, Manaster C, Teuber M, Rosenstiel P, Kwiatkowski R, Huse K *et al* (2006). SNPSplicer: systematic analysis of SNP-dependent splicing in genotyped cDNAs. *Hum Mutat* **27**: 1129-34.
- Fan JB, Oliphant A, Shen R, Kermani BG, Garcia F, Gunderson KL *et al* (2003). Highly parallel SNP genotyping. *Cold Spring Harb Symp Quant Biol* **68**: 69-78.
- Fei P, Yin J, Wang W (2005). New advances in the DNA damage response network of Fanconi anemia and BRCA proteins. FAAP95 replaces BRCA2 as the true FANCB protein. *Cell Cycle* **4**: 80-6.
- Fernandez-Capetillo O, Lee A, Nussenzweig M, Nussenzweig A (2004). H2AX: the histone guardian of the genome. *DNA Repair (Amst)* **3**: 959-67.
- Folias A, Matkovic M, Bruun D, Reid S, Hejna J, Grompe M *et al* (2002). BRCA1 interacts directly with the Fanconi anemia protein FANCA. *Hum Mol Genet* **11**: 2591-7.
- Freedman ML, Penney KL, Stram DO, Le Marchand L, Hirschhorn JN, Kolonel LN *et al* (2004). Common variation in BRCA2 and breast cancer risk: a haplotype-based analysis in the Multiethnic Cohort. *Hum Mol Genet* **13**: 2431-41.
- Gapska P, Scott RJ, Serrano-Fernandez P, Huzarski T, Byrski T, Kladny J *et al* (2008). Vitamin D receptor variants and breast cancer risk in the Polish population. *Breast Cancer Res Treat*.
- Garcia-Closas M, Hall P, Nevanlinna H, Pooley K, Morrison J, Richesson DA *et al* (2008). Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. *PLoS Genet* **4**: e1000054.
- Garcia-Higuera I, Taniguchi T, Ganesan S, Meyn MS, Timmers C, Hejna J *et al* (2001). Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway. *Mol Cell* **7**: 249-62.
- Garcia MJ, Benitez J (2008). The Fanconi anaemia/BRCA pathway and cancer susceptibility. Searching for new therapeutic targets. *Clin Transl Oncol* **10**: 78-84.

- Garcia MJ, Fernandez V, Osorio A, Barroso A, Lloret G, Lazaro C *et al* (2008). Analysis of FANCB and FANCN/PALB2 Fanconi Anemia genes in BRCA1/2-negative Spanish breast cancer families. *Breast Cancer Res Treat*.
- Garcia MJ, Fernandez V, Osorio A, Barroso A, Lloret G, Lazaro C *et al* (2009). Analysis of FANCB and FANCN/PALB2 fanconi anemia genes in BRCA1/2-negative Spanish breast cancer families. *Breast Cancer Res Treat* **113**: 545-51.
- Gatei M, Zhou BB, Hobson K, Scott S, Young D, Khanna KK (2001). Ataxia telangiectasia mutated (ATM) kinase and ATM and Rad3 related kinase mediate phosphorylation of Brca1 at distinct and overlapping sites. In vivo assessment using phospho-specific antibodies. *J Biol Chem* **276**: 17276-80.
- Georges M, Coppieters W, Charlier C (2007). Polymorphic miRNA-mediated gene regulation: contribution to phenotypic variation and disease. *Curr Opin Genet Dev* **17**: 166-76.
- Ghoussaini M, Song H, Koessler T, Al Olama AA, Kote-Jarai Z, Driver KE *et al* (2008). Multiple loci with different cancer specificities within the 8q24 gene desert. *J Natl Cancer Inst* **100**: 962-6.
- Gilad LA, Bresler T, Gnainsky J, Smirnoff P, Schwartz B (2005). Regulation of vitamin D receptor expression via estrogen-induced activation of the ERK 1/2 signaling pathway in colon and breast cancer cells. *J Endocrinol* **185**: 577-92.
- Gilad LA, Schwartz B (2007). Association of estrogen receptor beta with plasma-membrane caveola components: implication in control of vitamin D receptor. *J Mol Endocrinol* **38**: 603-18.
- Gordon SM, Alon N, Buchwald M (2005). FANCC, FANCE, and FANCD2 form a ternary complex essential to the integrity of the Fanconi anemia DNA damage response pathway. *J Biol Chem* **280**: 36118-25.
- Gordon SM, Buchwald M (2003). Fanconi anemia protein complex: mapping protein interactions in the yeast 2- and 3-hybrid systems. *Blood* **102**: 136-41.
- Greene FL, Sobin LH (2009). A worldwide approach to the TNM staging system: Collaborative efforts of the AJCC and UICC. *J Surg Oncol*.
- Grompe M (2002). FANCD2: a branch-point in DNA damage response? *Nat Med* **8**: 555-6.
- Grushko TA, Blackwood MA, Schumm PL, Hagos FG, Adeyanju MO, Feldman MD *et al* (2002). Molecular-cytogenetic analysis of HER-2/neu gene in BRCA1-associated breast cancers. *Cancer Res* **62**: 1481-8.
- Guy M, Lowe LC, Bretherton-Watt D, Mansi JL, Peckitt C, Bliss J *et al* (2004). Vitamin D receptor gene polymorphisms and breast cancer risk. *Clin Cancer Res* **10**: 5472-81.
- Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B *et al* (1990). Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* **250**: 1684-9.
- Hallstrom TC, Mori S, Nevins JR (2008). An E2F1-dependent gene expression program that determines the balance between proliferation and cell death. *Cancer Cell* **13**: 11-22.
- Han W, Kang D, Park IA, Kim SW, Bae JY, Chung KW *et al* (2004). Associations between breast cancer susceptibility gene polymorphisms and clinicopathological features. *Clin Cancer Res* **10**: 124-30.
- Hedenfalk I, Ringner M, Ben-Dor A, Yakhini Z, Chen Y, Chebil G *et al* (2003). Molecular classification of familial non-BRCA1/BRCA2 breast cancer. *Proc Natl Acad Sci U S A* **100**: 2532-7.

- Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG *et al* (2005). Whole-genome patterns of common DNA variation in three human populations. *Science* **307**: 1072-9.
- Hirschhorn JN, Daly MJ (2005). Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* **6**: 95-108.
- Holzel M, van Diest PJ, Bier P, Wallisch M, Hoatlin ME, Joenje H *et al* (2003). FANCD2 protein is expressed in proliferating cells of human tissues that are cancer-prone in Fanconi anaemia. *J Pathol* **201**: 198-203.
- Hong CC, Thompson HJ, Jiang C, Hammond GL, Tritchler D, Yaffe M *et al* (2003). Val158Met Polymorphism in catechol-O-methyltransferase gene associated with risk factors for breast cancer. *Cancer Epidemiol Biomarkers Prev* **12**: 838-47.
- Hou MF, Tien YC, Lin GT, Chen CJ, Liu CS, Lin SY *et al* (2002). Association of vitamin D receptor gene polymorphism with sporadic breast cancer in Taiwanese patients. *Breast Cancer Res Treat* **74**: 1-7.
- Houghtaling S, Granville L, Akkari Y, Torimaru Y, Olson S, Finegold M *et al* (2005). Heterozygosity for p53 (Trp53+/-) accelerates epithelial tumor formation in fanconi anemia complementation group D2 (Fancd2) knockout mice. *Cancer Res* **65**: 85-91.
- Houghtaling S, Timmers C, Noll M, Finegold MJ, Jones SN, Meyn MS *et al* (2003). Epithelial cancer in Fanconi anemia complementation group D2 (Fancd2) knockout mice. *Genes Dev* **17**: 2021-35.
- Howlett NG, Taniguchi T, Durkin SG, D'Andrea AD, Glover TW (2005). The Fanconi anemia pathway is required for the DNA replication stress response and for the regulation of common fragile site stability. *Hum Mol Genet* **14**: 693-701.
- Howlett NG, Taniguchi T, Olson S, Cox B, Waisfisz Q, De Die-Smulders C *et al* (2002). Biallelic inactivation of BRCA2 in Fanconi anemia. *Science* **297**: 606-9.
- Hull J, Campino S, Rowlands K, Chan MS, Copley RR, Taylor MS *et al* (2007). Identification of common genetic variation that modulates alternative splicing. *PLoS Genet* **3**: e99.
- Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE *et al* (2007). A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* **39**: 870-4.
- Hussain S, Wilson JB, Medhurst AL, Hejna J, Witt E, Ananth S *et al* (2004). Direct interaction of FANCD2 with BRCA2 in DNA damage response pathways. *Hum Mol Genet* **13**: 1241-8.
- Hussain S, Witt E, Huber PA, Medhurst AL, Ashworth A, Mathew CG (2003). Direct interaction of the Fanconi anaemia protein FANCG with BRCA2/FANCD1. *Hum Mol Genet* **12**: 2503-10.
- Ioannidis JP (2007). Non-replication and inconsistency in the genome-wide association setting. *Hum Hered* **64**: 203-13.
- Irvin WJ, Jr., Carey LA (2008). What is triple-negative breast cancer? *Eur J Cancer* **44**: 2799-805.
- Jacquemont C, Taniguchi T (2007). The Fanconi anemia pathway and ubiquitin. *BMC Biochem* **8 Suppl 1**: S10.
- Jasin M (2002). Homologous repair of DNA damage and tumorigenesis: the BRCA connection. *Oncogene* **21**: 8981-93.
- Joenje H, Oostra AB, Wijker M, di Summa FM, van Berkel CG, Rooimans MA *et al* (1997). Evidence for at least eight Fanconi anemia genes. *Am J Hum Genet* **61**: 940-4.
- John EM, Schwartz GG, Koo J, Wang W, Ingles SA (2007). Sun exposure, vitamin D receptor gene polymorphisms, and breast cancer risk in a multiethnic population. *Am J Epidemiol* **166**: 1409-19.

- Jurutka PW, Whitfield GK, Hsieh JC, Thompson PD, Haussler CA, Haussler MR (2001). Molecular nature of the vitamin D receptor and its role in regulation of gene expression. *Rev Endocr Metab Disord* **2**: 203-16.
- Kahali B, Basak S, Ghosh TC (2007). Reinvestigating the codon and amino acid usage of *S. cerevisiae* genome: a new insight from protein secondary structure analysis. *Biochem Biophys Res Commun* **354**: 693-9.
- Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E (2007). The role of site accessibility in microRNA target recognition. *Nat Genet* **39**: 1278-84.
- Kinzler KW, Vogelstein B (1997). Cancer-susceptibility genes. Gatekeepers and caretakers. *Nature* **386**: 761, 763.
- Kitao H, Yamamoto K, Matsushita N, Ohzeki M, Ishiai M, Takata M (2006). Functional interplay between BRCA2/FancD1 and FancC in DNA repair. *J Biol Chem* **281**: 21312-20.
- Kliewer EV, Smith KR (1995). Breast cancer mortality among immigrants in Australia and Canada. *J Natl Cancer Inst* **87**: 1154-61.
- Knight JA, Lesosky M, Barnett H, Raboud JM, Vieth R (2007). Vitamin D and reduced risk of breast cancer: a population-based case-control study. *Cancer Epidemiol Biomarkers Prev* **16**: 422-9.
- Koren M, Kimmel G, Ben-Asher E, Gal I, Papa MZ, Beckmann JS *et al* (2006). ATM haplotypes and breast cancer risk in Jewish high-risk women. *Br J Cancer* **94**: 1537-43.
- Lesueur F, Song H, Ahmed S, Luccarini C, Jordan C, Luben R *et al* (2006). Single-nucleotide polymorphisms in the RB1 gene and association with breast cancer in the British population. *Br J Cancer* **94**: 1921-6.
- Leveille F, Blom E, Medhurst AL, Bier P, Laghmani el H, Johnson M *et al* (2004). The Fanconi anemia gene product FANCF is a flexible adaptor protein. *J Biol Chem* **279**: 39421-30.
- Leveille F, Ferrer M, Medhurst AL, Laghmani el H, Rooimans MA, Bier P *et al* (2006). The nuclear accumulation of the Fanconi anemia protein FANCE depends on FANCC. *DNA Repair (Amst)* **5**: 556-65.
- Levitus M, Waisfisz Q, Godthelp BC, de Vries Y, Hussain S, Wiegant WW *et al* (2005). The DNA helicase BRIP1 is defective in Fanconi anemia complementation group J. *Nat Genet* **37**: 934-5.
- Levrán O, Attwooll C, Henry RT, Milton KL, Neveling K, Rio P *et al* (2005). The BRCA1-interacting helicase BRIP1 is deficient in Fanconi anemia. *Nat Genet* **37**: 931-3.
- Lewis BP, Burge CB, Bartel DP (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**: 15-20.
- Litman R, Peng M, Jin Z, Zhang F, Zhang J, Powell S *et al* (2005). BACH1 is critical for homologous recombination and appears to be the Fanconi anemia gene product FANCF. *Cancer Cell* **8**: 255-65.
- Livak KJ, Marmaro J, Todd JA (1995). Towards fully automated genome-wide polymorphism screening. *Nat Genet* **9**: 341-2.
- Lo YL, Yu JC, Chen ST, Hsu GC, Mau YC, Yang SL *et al* (2007). Breast cancer risk associated with genotypic polymorphism of the mitotic checkpoint genes: a multigenic study on cancer susceptibility. *Carcinogenesis* **28**: 1079-86.
- Loots GG, Ovcharenko I (2004). rVISTA 2.0: evolutionary analysis of transcription factor binding sites. *Nucleic Acids Res* **32**: W217-21.

- Lopez-Abente G, Pollan M, Vergara A, Ardanaz E, Moreo P, Moreno C *et al* (2000). [Time trends in cancer incidence in Navarra and Zaragoza, Spain]. *Gac Sanit* **14**: 100-9.
- Lyakhovich A, Surrallés J (2007). New roads to FA/BRCA pathway: H2AX. *Cell Cycle* **6**: 1019-23.
- Macaluso M, Cinti C, Russo G, Russo A, Giordano A (2003). pRb2/p130-E2F4/5-HDAC1-SUV39H1-p300 and pRb2/p130-E2F4/5-HDAC1-SUV39H1-DNMT1 multimolecular complexes mediate the transcription of estrogen receptor- α in breast cancer. *Oncogene* **22**: 3511-7.
- Maillet P, Bonnefoi H, Vaudan-Vutskits G, Pajk B, Cufer T, Foulkes WD *et al* (2002). Constitutional alterations of the ATM gene in early onset sporadic breast cancer. *J Med Genet* **39**: 751-3.
- Maiti B, Li J, de Bruin A, Gordon F, Timmers C, Opavsky R *et al* (2005). Cloning and characterization of mouse E2F8, a novel mammalian E2F family member capable of blocking cellular proliferation. *J Biol Chem* **280**: 18211-20.
- Marek LR, Bale AE (2006). Drosophila homologs of FANCD2 and FANCL function in DNA repair. *DNA Repair (Amst)* **5**: 1317-26.
- Marmorstein LY, Ouchi T, Aaronson SA (1998). The BRCA2 gene product functionally interacts with p53 and RAD51. *Proc Natl Acad Sci U S A* **95**: 13869-74.
- Masters JR (2002). HeLa cells 50 years on: the good, the bad and the ugly. *Nat Rev Cancer* **2**: 315-9.
- Mavaddat N, Dunning AM, Ponder BA, Easton DF, Pharoah PD (2009). Common genetic variation in candidate genes and susceptibility to subtypes of breast cancer. *Cancer Epidemiol Biomarkers Prev* **18**: 255-9.
- McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP *et al* (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* **9**: 356-69.
- McCauley JL, Kenealy SJ, Margulies EH, Schnetz-Boutaud N, Gregory SG, Hauser SL *et al* (2007). SNPs in Multi-species Conserved Sequences (MCS) as useful markers in association studies: a practical approach. *BMC Genomics* **8**: 266.
- McCullough ML, Stevens VL, Diver WR, Feigelson HS, Rodriguez C, Bostick RM *et al* (2007). Vitamin D pathway gene polymorphisms, diet, and risk of postmenopausal breast cancer: a nested case-control study. *Breast Cancer Res* **9**: R9.
- Medhurst AL, Huber PA, Waisfisz Q, de Winter JP, Mathew CG (2001). Direct interactions of the five known Fanconi anaemia proteins suggest a common functional pathway. *Hum Mol Genet* **10**: 423-9.
- Meetei AR, de Winter JP, Medhurst AL, Wallisch M, Waisfisz Q, van de Vrugt HJ *et al* (2003). A novel ubiquitin ligase is deficient in Fanconi anemia. *Nat Genet* **35**: 165-70.
- Meetei AR, Levitus M, Xue Y, Medhurst AL, Zwaan M, Ling C *et al* (2004a). X-linked inheritance of Fanconi anemia complementation group B. *Nat Genet* **36**: 1219-24.
- Meetei AR, Medhurst AL, Ling C, Xue Y, Singh TR, Bier P *et al* (2005). A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M. *Nat Genet* **37**: 958-63.
- Meetei AR, Yan Z, Wang W (2004b). FANCL replaces BRCA1 as the likely ubiquitin ligase responsible for FANCD2 monoubiquitination. *Cell Cycle* **3**: 179-81.
- Meijers-Heijboer H, van den Ouweland A, Klijn J, Wasielewski M, de Snoo A, Oldenburg R *et al* (2002). Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nat Genet* **31**: 55-9.

- Meyer-Rochow GY, Smith JM, Richardson AL, Marsh DJ, Sidhu SB, Robinson BG *et al* (2008). Denaturing High Performance Liquid Chromatography Detection of SDHB, SDHD, and VHL Germline Mutations in Pheochromocytoma. *J Surg Res*.
- Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S *et al* (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266**: 66-71.
- Milne RL, Osorio A, Ramón y Cajal T, Vega A, Lloret G, de la Hoya M *et al* (2008). The average cumulative risks of breast and ovarian cancer for carriers of mutations in BRCA1 and BRCA2 attending genetic counseling units in Spain. *Clinical Cancer Research* **14**: 2861-2869.
- Milne RL, Ribas G, Gonzalez-Neira A, Fagerholm R, Salas A, Gonzalez E *et al* (2006). ERCC4 associated with breast cancer risk: a two-stage case-control study using high-throughput genotyping. *Cancer Res* **66**: 9420-7.
- Mishra PJ, Banerjee D, Bertino JR (2008). MiRSNPs or MiR-polymorphisms, new players in microRNA mediated regulation of the cell: Introducing microRNA pharmacogenomics. *Cell Cycle* **7**: 853-8.
- Mishra PJ, Humeniuk R, Longo-Sorbello GS, Banerjee D, Bertino JR (2007). A miR-24 microRNA binding-site polymorphism in dihydrofolate reductase gene leads to methotrexate resistance. *Proc Natl Acad Sci U S A* **104**: 13513-8.
- Mitrunen K, Hirvonen A (2003). Molecular epidemiology of sporadic breast cancer. The role of polymorphic genes involved in oestrogen biosynthesis and metabolism. *Mutat Res* **544**: 9-41.
- Mitrunen K, Sillanpaa P, Kataja V, Eskelinen M, Kosma VM, Benhamou S *et al* (2001). Association between manganese superoxide dismutase (MnSOD) gene polymorphism and breast cancer risk. *Carcinogenesis* **22**: 827-9.
- Monteiro AN, Zhang S, Phelan CM, Narod SA (2003). Absence of constitutional H2AX gene mutations in 101 hereditary breast cancer families. *J Med Genet* **40**: e51.
- Montes de Oca R, Andreassen PR, Margossian SP, Gregory RC, Taniguchi T, Wang X *et al* (2005). Regulated interaction of the Fanconi anemia protein, FANCD2, with chromatin. *Blood* **105**: 1003-9.
- Morrison C, Sonoda E, Takao N, Shinohara A, Yamamoto K, Takeda S (2000). The controlling role of ATM in homologous recombinational repair of DNA damage. *Embo J* **19**: 463-71.
- Moynahan ME, Chiu JW, Koller BH, Jasin M (1999). Brca1 controls homology-directed DNA repair. *Mol Cell* **4**: 511-8.
- Moynahan ME, Pierce AJ, Jasin M (2001). BRCA2 is required for homology-directed repair of chromosomal breaks. *Mol Cell* **7**: 263-72.
- Nakanishi K, Moran A, Hays T, Kuang Y, Fox E, Garneau D *et al* (2001). Functional analysis of patient-derived mutations in the Fanconi anemia gene, FANCG/XRCC9. *Exp Hematol* **29**: 842-9.
- Narayan G, Arias-Pulido H, Nandula SV, Basso K, Sugirtharaj DD, Vargas H *et al* (2004). Promoter hypermethylation of FANCF: disruption of Fanconi Anemia-BRCA pathway in cervical cancer. *Cancer Res* **64**: 2994-7.
- Nielsen KB, Sorensen S, Cartegni L, Corydon TJ, Doktor TK, Schroeder LD *et al* (2007). Seemingly neutral polymorphic variants may confer immunity to splicing-inactivating mutations: a synonymous SNP in exon 5 of MCAD protects from deleterious mutations in a flanking exonic splicing enhancer. *Am J Hum Genet* **80**: 416-32.

- Nightingale KP, Gendreizig S, White DA, Bradbury C, Hollfelder F, Turner BM (2007). Cross-talk between histone modifications in response to histone deacetylase inhibitors: MLL4 links histone H3 acetylation and histone H3K4 methylation. *J Biol Chem* **282**: 4408-16.
- Nijman SM, Huang TT, Dirac AM, Brummelkamp TR, Kerkhoven RM, D'Andrea AD *et al* (2005). The deubiquitinating enzyme USP1 regulates the Fanconi anemia pathway. *Mol Cell* **17**: 331-9.
- Notredame C, Higgins DG, Heringa J (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302**: 205-17.
- O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT (2005). c-Myc-regulated microRNAs modulate E2F1 expression. *Nature* **435**: 839-43.
- Ohashi A, Zdzienicka MZ, Chen J, Couch FJ (2005). Fanconi anemia complementation group D2 (FANCD2) functions independently of BRCA2- and RAD51-associated homologous recombination in response to DNA damage. *J Biol Chem* **280**: 14877-83.
- Oldenburg RA, Meijers-Heijboer H, Cornelisse CJ, Devilee P (2007). Genetic susceptibility for breast cancer: how many more genes to be found? *Crit Rev Oncol Hematol* **63**: 125-49.
- Osorio A, Barroso A, Martinez B, Cebrian A, San Roman JM, Lobo F *et al* (2000). Molecular analysis of the BRCA1 and BRCA2 genes in 32 breast and/or ovarian cancer Spanish families. *Br J Cancer* **82**: 1266-70.
- Ouchi M, Fujiuchi N, Sasai K, Katayama H, Minamishima YA, Ongusaha PP *et al* (2004). BRCA1 phosphorylation by Aurora-A in the regulation of G2 to M transition. *J Biol Chem* **279**: 19643-8.
- Ovcharenko I, Nobrega MA, Loots GG, Stubbs L (2004). ECR Browser: a tool for visualizing and accessing data from comparisons of multiple vertebrate genomes. *Nucleic Acids Res* **32**: W280-6.
- Palacios J, Robles-Frias MJ, Castilla MA, Lopez-Garcia MA, Benitez J (2008). The molecular pathology of hereditary breast cancer. *Pathobiology* **75**: 85-94.
- Park WH, Margossian S, Horwitz AA, Simons AM, D'Andrea AD, Parvin JD (2005). Direct DNA binding activity of the Fanconi anemia D2 protein. *J Biol Chem* **280**: 23593-8.
- Parkin DM, Bray F, Ferlay J, Pisani P (2005). Global cancer statistics, 2002. *CA Cancer J Clin* **55**: 74-108.
- Peters L, Meister G (2007). Argonaute proteins: mediators of RNA silencing. *Mol Cell* **26**: 611-23.
- Pharoah PD, Antoniou A, Bobrow M, Zimmern RL, Easton DF, Ponder BA (2002). Polygenic susceptibility to breast cancer and implications for prevention. *Nat Genet* **31**: 33-6.
- Pharoah PD, Antoniou AC, Easton DF, Ponder BA (2008). Polygenes, risk prediction, and targeted prevention of breast cancer. *N Engl J Med* **358**: 2796-803.
- Pharoah PD, Dunning AM, Ponder BA, Easton DF (2004). Association studies for finding cancer-susceptibility genetic variants. *Nat Rev Cancer* **4**: 850-60.
- Pharoah PD, Tyrer J, Dunning AM, Easton DF, Ponder BA (2007). Association between common variation in 120 candidate genes and breast cancer risk. *PLoS Genet* **3**: e42.
- Polager S, Ginsberg D (2008). E2F - at the crossroads of life and death. *Trends Cell Biol* **18**: 528-35.
- Polager S, Kalma Y, Berkovich E, Ginsberg D (2002). E2Fs up-regulate expression of genes involved in DNA replication, DNA repair and mitosis. *Oncogene* **21**: 437-46.
- Ponder BA (2001). Cancer genetics. *Nature* **411**: 336-41.

- Pooley KA, Baynes C, Driver KE, Tyrer J, Azzato EM, Pharoah PD *et al* (2008). Common single-nucleotide polymorphisms in DNA double-strand break repair genes and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* **17**: 3482-9.
- Powell SN, Willers H, Xia F (2002). BRCA2 keeps Rad51 in line. High-fidelity homologous recombination prevents breast and ovarian cancer? *Mol Cell* **10**: 1262-3.
- Powers JT, Hong S, Mayhew CN, Rogers PM, Knudsen ES, Johnson DG (2004). E2F1 uses the ATM signaling pathway to induce p53 and Chk2 phosphorylation and apoptosis. *Mol Cancer Res* **2**: 203-14.
- Prokunina L, Alarcon-Riquelme ME (2004). Regulatory SNPs in complex diseases: their identification and functional validation. *Expert Rev Mol Med* **6**: 1-15.
- Pylkas K, Tommiska J, Syrjakoski K, Kere J, Gatei M, Waddell N *et al* (2007). Evaluation of the role of Finnish ataxia-telangiectasia mutations in hereditary predisposition to breast cancer. *Carcinogenesis* **28**: 1040-5.
- Rahman N, Ashworth A (2004). A new gene on the X involved in Fanconi anemia. *Nat Genet* **36**: 1142-3.
- Rahman N, Seal S, Thompson D, Kelly P, Renwick A, Elliott A *et al* (2007). PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nat Genet* **39**: 165-7.
- Renwick A, Thompson D, Seal S, Kelly P, Chagtai T, Ahmed M *et al* (2006). ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nat Genet* **38**: 873-5.
- Risch N (2001). The genetic epidemiology of cancer: interpreting family and twin studies and their implications for molecular genetic approaches. *Cancer Epidemiol Biomarkers Prev* **10**: 733-41.
- Risch NJ (2000). Searching for genetic determinants in the new millennium. *Nature* **405**: 847-56.
- Rodriguez-Lopez R, Osorio A, Sanchez-Pulido L, De La Hoya M, Barroso A, Caldes T *et al* (2003). No mutations in the XRCC2 gene in BRCA1/2-negative high-risk breast cancer families. *Int J Cancer* **103**: 136-7.
- Rogoff HA, Kowalik TF (2004). Life, death and E2F: linking proliferation control and DNA damage signaling via E2F1. *Cell Cycle* **3**: 845-6.
- Rokavec M, Justenhoven C, Schroth W, Istrate MA, Haas S, Fischer HP *et al* (2007). A novel polymorphism in the promoter region of ERBB4 is associated with breast and colorectal cancer risk. *Clin Cancer Res* **13**: 7506-14.
- Rosa-Rosa JM, Pita G, Urioste M, Llorca G, Brunet J, Lazaro C *et al* (2009). Genome-wide linkage scan reveals three putative breast-cancer-susceptibility loci. *Am J Hum Genet* **84**: 115-22.
- Rossner P, Jr., Terry MB, Gammon MD, Zhang FF, Teitelbaum SL, Eng SM *et al* (2006). OGG1 polymorphisms and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* **15**: 811-5.
- Rothfuss A, Grompe M (2004). Repair kinetics of genomic interstrand DNA cross-links: evidence for DNA double-strand break-dependent activation of the Fanconi anemia/BRCA pathway. *Mol Cell Biol* **24**: 123-34.
- Rozen S, Skaletsky H (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **132**: 365-86.
- Russo J, Hu YF, Yang X, Russo IH (2000). Developmental, cellular, and molecular basis of human breast cancer. *J Natl Cancer Inst Monogr*: 17-37.
- Rylander-Rudqvist T, Wedren S, Granath F, Humphreys K, Ahlberg S, Weiderpass E *et al* (2003). Cytochrome P450 1B1 gene polymorphisms and postmenopausal breast cancer risk. *Carcinogenesis* **24**: 1533-9.

- Saunders MA, Liang H, Li WH (2007). Human polymorphism at microRNAs and microRNA target sites. *Proc Natl Acad Sci U S A* **104**: 3300-5.
- Sazci A, Ergul E, Utkan NZ, Canturk NZ, Kaya G (2004). Catechol-O-methyltransferase Val 108/158 Met polymorphism in premenopausal breast cancer patients. *Toxicology* **204**: 197-202.
- Schattner P, Diekhans M (2006). Regions of extreme synonymous codon selection in mammalian genes. *Nucleic Acids Res* **34**: 1700-10.
- Scheijen B, Bronk M, van der Meer T, De Jong D, Bernards R (2004). High incidence of thymic epithelial tumors in E2F2 transgenic mice. *J Biol Chem* **279**: 10476-83.
- Schrauder M, Frank S, Strissel PL, Lux MP, Bani MR, Rauh C *et al* (2008). Single nucleotide polymorphism D1853N of the ATM gene may alter the risk for breast cancer. *J Cancer Res Clin Oncol* **134**: 873-82.
- Scintu M, Vitale R, Prencipe M, Gallo AP, Bonghi L, Valori VM *et al* (2007). Genomic instability and increased expression of BUB1B and MAD2L1 genes in ductal breast carcinoma. *Cancer Lett* **254**: 298-307.
- Scully R, Livingston DM (2000). In search of the tumour-suppressor functions of BRCA1 and BRCA2. *Nature* **408**: 429-32.
- Seal S, Thompson D, Renwick A, Elliott A, Kelly P, Barfoot R *et al* (2006). Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nat Genet* **38**: 1239-41.
- Shabalina SA, Ogurtsov AY, Spiridonov NA (2006). A periodic pattern of mRNA secondary structure created by the genetic code. *Nucleic Acids Res* **34**: 2428-37.
- Shen R, Fan JB, Campbell D, Chang W, Chen J, Doucet D *et al* (2005). High-throughput SNP genotyping on universal bead arrays. *Mutat Res* **573**: 70-82.
- Simin K, Wu H, Lu L, Pinkel D, Albertson D, Cardiff RD *et al* (2004). pRb inactivation in mammary cells reveals common mechanisms for tumor initiation and progression in divergent epithelia. *PLoS Biol* **2**: E22.
- Sims AE, Spiteri E, Sims RJ, 3rd, Arita AG, Lach FP, Landers T *et al* (2007). FANCI is a second monoubiquitinated member of the Fanconi anemia pathway. *Nat Struct Mol Biol* **14**: 564-7.
- Smith TR, Miller MS, Lohman K, Lange EM, Case LD, Mohrenweiser HW *et al* (2003). Polymorphisms of XRCC1 and XRCC3 genes and susceptibility to breast cancer. *Cancer Lett* **190**: 183-90.
- Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H *et al* (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* **98**: 10869-74.
- Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A *et al* (2003). Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* **100**: 8418-23.
- Soule HD, Vazquez J, Long A, Albert S, Brennan M (1973). A human cell line from a pleural effusion derived from a breast carcinoma. *J Natl Cancer Inst* **51**: 1409-16.
- Sprague BL, Trentham-Dietz A, Garcia-Closas M, Newcomb PA, Titus-Ernstoff L, Hampton JM *et al* (2007). Genetic variation in TP53 and risk of breast cancer in a population-based case control study. *Carcinogenesis* **28**: 1680-6.

- Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA *et al* (2007). Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* **39**: 865-9.
- Stewart G, Elledge SJ (2002). The two faces of BRCA2, a FANCTastic discovery. *Mol Cell* **10**: 2-4.
- Stredrick DL, Garcia-Closas M, Pineda MA, Bhatti P, Alexander BH, Doody MM *et al* (2006). The ATM missense mutation p.Ser49Cys (c.146C>G) and the risk of breast cancer. *Hum Mutat* **27**: 538-44.
- Sun T, Gao Y, Tan W, Ma S, Shi Y, Yao J *et al* (2007). A six-nucleotide insertion-deletion polymorphism in the CASP8 promoter is associated with susceptibility to multiple cancers. *Nat Genet* **39**: 605-13.
- Sun T, Miao X, Wang J, Tan W, Zhou Y, Yu C *et al* (2004). Functional Phe31Ile polymorphism in Aurora A and risk of breast carcinoma. *Carcinogenesis* **25**: 2225-30.
- Tamimi RM, Hankinson SE, Spiegelman D, Kraft P, Colditz GA, Hunter DJ (2004). Common ataxia telangiectasia mutated haplotypes and risk of breast cancer: a nested case-control study. *Breast Cancer Res* **6**: R416-22.
- Taniguchi T, Garcia-Higuera I, Andreassen PR, Gregory RC, Grompe M, D'Andrea AD (2002a). S-phase-specific interaction of the Fanconi anemia protein, FANCD2, with BRCA1 and RAD51. *Blood* **100**: 2414-20.
- Taniguchi T, Garcia-Higuera I, Xu B, Andreassen PR, Gregory RC, Kim ST *et al* (2002b). Convergence of the fanconi anemia and ataxia telangiectasia signaling pathways. *Cell* **109**: 459-72.
- Taron M, Rosell R, Felip E, Mendez P, Souglakos J, Ronco MS *et al* (2004). BRCA1 mRNA expression levels as an indicator of chemoresistance in lung cancer. *Hum Mol Genet* **13**: 2443-9.
- Taubert S, Gorrini C, Frank SR, Parisi T, Fuchs M, Chan HM *et al* (2004). E2F-dependent histone acetylation and recruitment of the Tip60 acetyltransferase complex to chromatin in late G1. *Mol Cell Biol* **24**: 4546-56.
- Tavtigian SV, Simard J, Rommens J, Couch F, Shattuck-Eidens D, Neuhausen S *et al* (1996). The complete BRCA2 gene and mutations in chromosome 13q-linked kindreds. *Nat Genet* **12**: 333-7.
- Thompson D, Easton D (2004). The genetic epidemiology of breast cancer genes. *J Mammary Gland Biol Neoplasia* **9**: 221-36.
- Thompson D, Szabo CI, Mangion J, Oldenburg RA, Odefrey F, Seal S *et al* (2002). Evaluation of linkage of breast cancer to the putative BRCA3 locus on chromosome 13q21 in 128 multiple case families from the Breast Cancer Linkage Consortium. *Proc Natl Acad Sci U S A* **99**: 827-31.
- Thompson E, Dragovic RL, Stephenson SA, Eccles DM, Campbell IG, Dobrovic A (2005). A novel duplication polymorphism in the FANCA promoter and its association with breast and ovarian cancer. *BMC Cancer* **5**: 43.
- Thorstenson YR, Roxas A, Kroiss R, Jenkins MA, Yu KM, Bachrich T *et al* (2003). Contributions of ATM mutations to familial breast and ovarian cancer. *Cancer Res* **63**: 3325-33.
- Timmers C, Sharma N, Opavsky R, Maiti B, Wu L, Wu J *et al* (2007). E2f1, E2f2, and E2f3 control E2F target expression and cellular proliferation via a p53-dependent negative feedback loop. *Mol Cell Biol* **27**: 65-78.

- Tischkowitz M, Xia B, Sabbaghian N, Reis-Filho JS, Hamel N, Li G *et al* (2007). Analysis of PALB2/FANCD1-associated breast cancer families. *Proc Natl Acad Sci U S A* **104**: 6788-93.
- Trabert B, Malone KE, Daling JR, Doody DR, Bernstein L, Ursin G *et al* (2007). Vitamin D receptor polymorphisms and breast cancer risk in a large population-based case-control study of Caucasian and African-American women. *Breast Cancer Res* **9**: R84.
- Trang HM, Cole DE, Rubin LA, Pierratos A, Siu S, Vieth R (1998). Evidence that vitamin D3 increases serum 25-hydroxyvitamin D more efficiently than does vitamin D2. *Am J Clin Nutr* **68**: 854-8.
- Vaclavicek A, Bermejo JL, Wappenschmidt B, Meindl A, Sutter C, Schmutzler RK *et al* (2007). Genetic variation in the major mitotic checkpoint genes does not affect familial breast cancer risk. *Breast Cancer Res Treat* **106**: 205-13.
- van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M *et al* (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**: 530-6.
- van den Heuvel S, Dyson NJ (2008). Conserved functions of the pRB and E2F families. *Nat Rev Mol Cell Biol* **9**: 713-24.
- van der Groep P, Hoelzel M, Buerger H, Joenje H, de Winter JP, van Diest PJ (2008). Loss of expression of FANCD2 protein in sporadic and hereditary breast cancer. *Breast Cancer Res Treat* **107**: 41-7.
- van der Heijden MS, Yeo CJ, Hruban RH, Kern SE (2003). Fanconi anemia gene mutations in young-onset pancreatic cancer. *Cancer Res* **63**: 2585-8.
- Venkitaraman AR (2002). Connecting Fanconi's anaemia to breast cancer predisposition. *Lancet* **360**: 1344-5.
- Venkitaraman AR (2004). Tracing the network connecting BRCA and Fanconi anaemia proteins. *Nat Rev Cancer* **4**: 266-76.
- Vogelstein B, Kinzler KW (2004). Cancer genes and the pathways they control. *Nat Med* **10**: 789-99.
- Vuaroqueaux V, Urban P, Labuhn M, Delorenzi M, Wirapati P, Benz CC *et al* (2007). Low E2F1 transcript levels are a strong determinant of favorable breast cancer outcome. *Breast Cancer Res* **9**: R33.
- Walsh T, Casadei S, Coats KH, Swisher E, Stray SM, Higgins J *et al* (2006). Spectrum of mutations in BRCA1, BRCA2, CHEK2, and TP53 in families at high risk of breast cancer. *Jama* **295**: 1379-88.
- Wang A, Schneider-Broussard R, Kumar AP, MacLeod MC, Johnson DG (2000a). Regulation of BRCA1 expression by the Rb-E2F pathway. *J Biol Chem* **275**: 4532-6.
- Wang S, Zhang B, Faller DV (2004a). BRG1/BRM and prohibitin are required for growth suppression by estrogen antagonists. *EMBO J* **23**: 2293-303.
- Wang WY, Barratt BJ, Clayton DG, Todd JA (2005). Genome-wide association studies: theoretical and practical concerns. *Nat Rev Genet* **6**: 109-18.
- Wang X, Andreassen PR, D'Andrea AD (2004b). Functional interaction of monoubiquitinated FANCD2 and BRCA2/FANCD1 in chromatin. *Mol Cell Biol* **24**: 5850-62.
- Wang X, D'Andrea AD (2004). The interplay of Fanconi anemia proteins in the DNA damage response. *DNA Repair (Amst)* **3**: 1063-9.
- Wang Y, Cortez D, Yazdi P, Neff N, Elledge SJ, Qin J (2000b). BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures. *Genes Dev* **14**: 927-39.

- Wang Z, Li M, Lu S, Zhang Y, Wang H (2006). Promoter hypermethylation of FANCF plays an important role in the occurrence of ovarian cancer through disrupting Fanconi anemia-BRCA pathway. *Cancer Biol Ther* **5**: 256-60.
- Wedren S, Rudqvist TR, Granath F, Weiderpass E, Ingelman-Sundberg M, Persson I *et al* (2003). Catechol-O-methyltransferase gene polymorphism and post-menopausal breast cancer risk. *Carcinogenesis* **24**: 681-7.
- Weintraub SJ, Prater CA, Dean DC (1992). Retinoblastoma protein switches the E2F site from positive to negative element. *Nature* **358**: 259-61.
- Wong AK, Shanahan F, Chen Y, Lian L, Ha P, Hendricks K *et al* (2000). BRG1, a component of the SWI-SNF complex, is mutated in multiple human tumor cell lines. *Cancer Res* **60**: 6171-7.
- Wood RD, Mitchell M, Sgouros J, Lindahl T (2001). Human DNA repair genes. *Science* **291**: 1284-9.
- Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J *et al* (1995). Identification of the breast cancer susceptibility gene BRCA2. *Nature* **378**: 789-92.
- Wooster R, Neuhausen SL, Mangion J, Quirk Y, Ford D, Collins N *et al* (1994). Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science* **265**: 2088-90.
- Wu L, Timmers C, Maiti B, Saavedra HI, Sang L, Chong GT *et al* (2001). The E2F1-3 transcription factors are essential for cellular proliferation. *Nature* **414**: 457-62.
- Yang J, Yu Y, Hamrick HE, Duerksen-Hughes PJ (2003). ATM, ATR and DNA-PK: initiators of the cellular genotoxic stress responses. *Carcinogenesis* **24**: 1571-80.
- Ye C, Dai Q, Lu W, Cai Q, Zheng Y, Shu XO *et al* (2007). Two-stage case-control study of common ATM gene variants in relation to breast cancer risk. *Breast Cancer Res Treat* **106**: 121-6.
- Yoshida K, Miki Y (2004). Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage. *Cancer Sci* **95**: 866-71.
- Yu Z, Li Z, Jolicoeur N, Zhang L, Fortin Y, Wang E *et al* (2007). Aberrant allele frequencies of the SNPs located in microRNA target sites are potentially associated with human cancers. *Nucleic Acids Res* **35**: 4535-41.
- Yuan B, Xu Y, Woo JH, Wang Y, Bae YK, Yoon DS *et al* (2006). Increased expression of mitotic checkpoint genes in breast cancer cells with chromosomal instability. *Clin Cancer Res* **12**: 405-10.
- Yuan SS, Lee SY, Chen G, Song M, Tomlinson GE, Lee EY (1999). BRCA2 is required for ionizing radiation-induced assembly of Rad51 complex in vivo. *Cancer Res* **59**: 3547-51.
- Zacharatos P, Kotsinas A, Evangelou K, Karakaidos P, Vassiliou LV, Rezaei N *et al* (2004). Distinct expression patterns of the transcription factor E2F-1 in relation to tumour growth parameters in common human carcinomas. *J Pathol* **203**: 744-53.
- Zhang SY, Liu SC, Al-Saleem LF, Holloran D, Babb J, Guo X *et al* (2000). E2F-1: a proliferative marker of breast neoplasia. *Cancer Epidemiol Biomarkers Prev* **9**: 395-401.
- Zhang Y, Newcomb PA, Egan KM, Titus-Ernstoff L, Chanock S, Welch R *et al* (2006). Genetic polymorphisms in base-excision repair pathway genes and risk of breast cancer. *Cancer Epidemiol Biomarkers Prev* **15**: 353-8.
- Zhu S, Wu H, Wu F, Nie D, Sheng S, Mo YY (2008). MicroRNA-21 targets tumor suppressor genes in invasion and metastasis. *Cell Res* **18**: 350-9.

VIII. ANEXOS

Anexo I.1. Descripción de los estadios establecidos según la codificación TNM. Codificación establecida por AJCC (American Joint Committee on Cancer) y UICC (Internacional Union Against Cancer).

Estadio	Descripción
Tumor primario	
TX	Tumor primario no valorable
T0	No hay evidencia de tumor primario
Tis	Carcinoma <i>in situ</i>
T1	Tumor de hasta 2 cm en su diámetro mayor
T2	Tumor >2 cm y <5 cm en su diámetro mayor
T3	Tumor >5 cm en su diámetro mayor
T4	Tumor de cualquier tamaño con extensión a la pared torácica o a la piel
Nódulos linfáticos regionales	
NX	Nódulos linfáticos regionales no valorables
N0	No hay evidencia de nódulos linfáticos regionales involucrados
N1	Metástasis a ganglios axilares ipsilaterales móviles
N2	Metástasis a ganglios axilares ipsilaterales fijos unos a otros o a otras estructuras
N3	Metástasis a ganglios linfáticos mamarios internos ipsilaterales
Metástasis distantes	
MX	Metástasis distantes no valorables
M0	No hay evidencia de metástasis distantes
M1	Existencia de metástasis distantes (incluyendo ganglios supraclaviculares ipsilaterales)

Anexo I.2. Distribución de muestras CoCM, CME y CMA por hospital o centro donador en toda la colección de muestras de ADN disponibles en el Grupo de Genética Humana.

Hospital/Centro	CoCM		CME		CMA		Total	
	N	%	N	%	N	%	N	%
Instituto Palacios	463	30,3	-	-	-	-	463	15,1
Colegio de Abogados	288	18,9	-	-	-	-	288	9,4
Centro Nacional de Transfusiones	287	18,8	-	-	-	-	287	9,4
Hospital Clínico de Barcelona	63	4,1	-	-	-	-	63	2,1
Instituto Catalán de Oncología	57	3,7	-	-	-	-	57	1,9
Centro de Investigación del Cáncer	13	0,9	-	-	-	-	13	0,4
Hospital Monte Naranco	-	-	389	30,9	5	1,8	394	12,9
Hospital La Paz	-	-	270	21,4	1	0,4	271	8,9
Fundación Jiménez Díaz	176	11,5	238	18,9	1	0,4	239	7,8
Hospital San Pau	-	-	115	9,1	-	-	115	3,8
Hospital Clínico San Carlos	-	-	102	8,1	-	-	102	3,3
Hospital de Salamanca	-	-	100	7,9	-	-	100	3,3
Hospital de Valencia	-	-	28	2,2	-	-	28	0,9
Desconocido	655	42,9	677	53,7	275	99,6	1.607	52,5
Total	1.526	100,0	1.260	100,0	276	100,0	3.062	100,0

CoCM, muestras control para cáncer de mama, CME, casos de cáncer de mama esporádico, y CMA, casos de cáncer de mama de agregación familiar.

Anexo I.3. A, Hoja 1 del modelo de cuestionario utilizado para la recolección de datos personales, clínicos y tumorales de interés para el estudio del cáncer de mama.

A. FORMULARIO DE RECOGIDA DE DATOS PARA MUJERES CON CÁNCER DE MAMA

Fecha Centro Código CNIO

dd mm yyyy

Nombre Apellido 1 Apellido 2

Fecha Nacimiento Edad Teléfono

dd mm yyyy

Lugar/País Nacimiento Edad de diagnóstico
(1^{er} cáncer de mama)

Lugar/País Nacimiento Madre

Lugar/País Nacimiento Padre

Altura (cm) Peso 1 año antes del diagnóstico (kg)

Edad Menarquia

Número de Partos Edad Primer Parto

Número de Abortos

Edad Menopausia

Causa Menopausia: Natural ☐ Cirugía ☐ Tratamiento ☐ Otros ☐

Consumo de Alcohol Habitual

☐ Nunca

☐ Menos de una copa/caña por semana

☐ Menos de una copa/caña al día

☐ 1-2 copas/cañas al día

☐ Más de 2 copas/cañas al día

Edad Inicio Duración (años)

Uso de Tabaco

☐ No (no-fumadora)

☐ No (ex-fumadora) Edad Final

☐ Sólo fuma ocasionalmente

☐ Sí (fumadora)

Cigarrillos fumados al día

Edad Inicio Duración (años)

Toma de Anticonceptivos: No ☐ Sí ☐ Edad Inicio Duración (años)

Estimulación Ovárica: No ☐ Sí ☐ Edad Inicio Número de ciclos

Tratamiento Hormonal Sustitutivo: No ☐ Sí ☐ Edad Inicio Duración (años)

Mamografías: ¿Cuántas ha tenido usted en su vida? Edad Primer Mamografía

¿Ha sido usted diagnosticada de cáncer previamente? No ☐ Sí ☐ Tipo

¿Tiene usted algún pariente de primer grado (madre, padre, hermano/a, hijo/a) que ha sido diagnosticado con cáncer de mama? No ☐ Sí ☐ Número

Continuación de Anexo I.3. B, Hoja 2 del modelo de cuestionario utilizado para la recolección de datos personales, clínicos y tumorales de interés para el estudio del cáncer de mama.

Datos recogidos por el médico oncólogo colaborador en las consultas de revisión de las pacientes.
B.

A RELLENAR POR EL MÉDICO	
Fecha de diagnóstico (1 ^{er} cáncer de mama) <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> dd mm yy	
Cirugía Mama	No <input type="checkbox"/> Sí <input type="checkbox"/>
Fecha de intervención <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> dd mm yy	Tipo de intervención <input type="text"/>
Hospital en que se realizó <input type="text"/>	
HISTOPATOLOGÍA	
Tipo histológico <input type="text"/>	
Grado de diferenciación <input type="checkbox"/>	
T <input type="checkbox"/>	N <input type="checkbox"/> M <input type="checkbox"/>
Inmunohistoquímica: RE <input type="checkbox"/> RP <input type="checkbox"/> Ki-67 <input type="checkbox"/> P53 <input type="checkbox"/> Erb2 <input type="checkbox"/>	
QUIMIOTERAPIA No <input type="checkbox"/> Sí <input type="checkbox"/>	
Fármacos utilizados <input type="text"/>	
Número de Ciclos <input type="checkbox"/>	<input type="text"/>
Toxicidades No <input type="checkbox"/> Sí <input type="checkbox"/>	Tipo <input type="text"/>
Respuesta al tratamiento No <input type="checkbox"/> Sí <input type="checkbox"/>	
Recidiva No <input type="checkbox"/> Si <input type="checkbox"/>	
Tiempo hasta recidiva desde finalización del tratamiento de quimioterapia <input type="text"/>	
TERAPIA HORMONAL No <input type="checkbox"/> Sí <input type="checkbox"/>	
Fármacos utilizados <input type="text"/>	
Toxicidades No <input type="checkbox"/> Sí <input type="checkbox"/>	Tipo <input type="text"/>
Recidiva No <input type="checkbox"/> Si <input type="checkbox"/>	
Tiempo hasta recidiva desde finalización del tratamiento de quimioterapia <input type="text"/>	

VIII. ANEXOS
VIII.1. Anexo I: Información adicional

Anexo I.4. Modelo de formulario utilizado para la recolección de datos personales, clínicos y tumorales de interés para el estudio del cáncer de mama.

A.

DNA ID Key	DNA ID Original	Other DNA ID	Bilateral BC	Surgery_1	Surgery Date_1	Surgery Hospital_1
Name	Initials	Sex	Surgery Type_1	Tumor Type_1	Tumor Specificities_1	
Person Type	Disease Type	N Family	Associated CDIS_1	Associated CDIS Specificities_1		
Reception Date	Diagnosis Date	Diagnosis Age	Tumor Position_1	Breast Position_1		
Hospital	N History File	Education	Differentiation_1	Chemotherapy_1		
Birthday	Birth Place	Foreigner	Histologic Grade_1	Neoadjuvance Chemotherapy Type_1		
Menarchy Age	Personal History of Cancer		Tumor Size_1	Adjuvance Chemotherapy Type_1		
Menopause Status	Personal History of Breast Cancer		N Positive Nodes / N Total Nodes_1	Chemotherapy Toxicity_1	Chemotherapy Toxicity Type_1	
Menopause Age	Other Personal History		Metastasis Status_1	Hormonal Therapy_1		
Menopause Cause	Family History of Cancer		TNM_1	Hormonal Therapy Type_1		
N Pregnancies	Family History of Breast Cancer 1*		ER_1	Hormonal Therapy Toxicity_1	Hormonal Therapy Toxicity Type_1	
N Aborts	Family History of Breast Cancer 2*		PR_1	Radiotherapy_1		
N Children	Family History of Other Cancer 1*		AR_1	Radiotherapy Type_1	Radiotherapy Location_1	
Age 1st Labor	Family History of Other Cancer 2*		Ki-67_1	Radiotherapy Toxicity_1	Radiotherapy Toxicity Type_1	
Lactancy			P53_1	Response to Treatment_1		
Contraceptives	Smoker	Life Status	Bcl2_1	Metastasis_1		
Age 1st Contraceptive	Ex-smoker	Date of Last Consultation	Erb2_1	1Metastasis Type_1	1Metastasis Date_1	
Years Contraceptive	Smoker Frequency	Monitoring Months	EGFR_1	1Months to Metastasis_1	1Treatment in Metastasis_1	
Ovaric Stimulation	N Cigarettes	Death Date	CatepsinaD_1	2Metastasis Type_1	2Metastasis Date_1	
Years Ovaric Stimulation	Age 1st Cigarette	Death Cause	CadherinaE_1	2Months to Metastasis_1	2Treatment in Metastasis_1	
THS	Years Smoking		PCNA_1	3Metastasis Type_1	3Metastasis Date_1	
Years THS Treatment	Alcohol		HER2/neu_1	3Months to Metastasis_1	3Treatment in Metastasis_1	
Mammography	Height		TPA_1	4Metastasis Type_1	4Metastasis Date_1	
N Mammography	Weight		pS2_1	4Months to Metastasis_1	4Treatment in Metastasis_1	
Age 1st Mammography	Obesity		Relapse_1	Response to Metastasis Treatment_1		
			Treatment in Relapse_1	Relapse Date_1	Months to Relapse_1	
				Metastasis in Relapse_1		

B.

DNA ID Key	Menarchy	Personal History of Cancer	Smoker
DNA ID Original	N Pregnancies	Personal History of Breast Cancer	Smoker Frequency
Person Type	N Aborts	Other Personal History	N Cigarettes
Reception Date	N Children	Family History of Cancer	Age 1st Cigarette
Diagnosis Age	Menopause Age		Years Smoking
Hospital	Menopause Status		Ex-smoker
Sex	Menopause Cause		Alcohol
Birthday	Contraceptives		Height
Birth Place	Age 1st Contraceptive		Weight
Foreigner	Years Contraceptive Treatment		Obesity
	Ovaric Stimulation		

Datos recogidos mediante la revisión de la historia clínica de las pacientes (A) o una entrevista personal con los controles (B).

Anexo I.5. Descripción de los oligonucleótidos utilizados en PCR y RT-Q-PCR.

Gen	Fragmento	pb	Oligonucleótidos		Tm (°C)	Condiciones de PCR (°C - segundos)			Condiciones de dHPLC (°C)	
			Sentido	Secuencia 5'→3'		Desnaturalización	Anillamiento	Elongación	Tm	Gradiente
FAND2	Promotor parte 1	392	Directo	CAGGCTGGACCAAAAGGT	64,0	95,0 - 15	63,8 - 15	72,0 - 45	60,5	58,9 - 68,9
			Reverso	ACTCTGAGCGCAGGGTCAT	65,1					
	Promotor parte 2	396	Directo	GCTGCATGCTATGGGGTATAG	63,2	95,0 - 15	63,8 - 15	72,0 - 45	56,3	59,0 - 69,0
			Reverso	GTTGGGAAGAGTGCTGGT	64,1					
	Promotor parte 3	399	Directo	TCACACTCACTTGAAGTGT	63,5	95,0 - 15	53,2 - 15	72,0 - 45	57,9	59,0 - 69,0
			Reverso	CTTCCCGAGCTACCTCAC	63,6					
	Promotor parte 4	381	Directo	AAGCCCTGCACTTCTTGTG	64,3	95,0 - 15	59,6 - 15	72,0 - 45	62,2	58,7 - 68,7
			Reverso	AGCCCGCTGAAGGCAATA	64,7					
	Exón 1	373	Directo	GCGCAGTTGCTTCTCTG	64,1	95,0 - 15	53,2 - 15	72,0 - 45	64,4	58,5 - 68,5
			Reverso	TATCCATCTCAGGGCAGAT	63,3					
	Exón 9	227	Directo	CTGGGTAATGTGCTGAGTT	62,7	95,0 - 15	56,0 - 15	72,0 - 45	-	-
			Reverso	TGCAGTGATAGAACCCCATTA	64,1					
	Exón 16	300	Directo	CCAAAAGGATAGCAACTTTAGG	62,8	95,0 - 15	60,0-55,0 - 15	72,0 - 45	56,7	56,7 - 66,7
			Reverso	TGAGTTCCAAACCAACAGAAA	62,0					
	Exón 19	194	Directo	CCTCTTTTGTGTGCCATT	64,0	95,0 - 15	56,0 - 15	72,0 - 45	-	-
			Reverso	CCACAGAAGTCAGAACTCCA	61,8					
	Exón 23	244	Directo	CGTATCTGAGCTGCAACA	60,0	95,0 - 15	56,0-53,0 - 15	72,0 - 45	57,8	56,7 - 66,7
	L714P		Reverso	CCCCATGAACATTCAAAACG	60,0					
	Exón 38	243	Directo	TTGGCTGTGACTCAGAGGTG	64,2	95,0 - 5	63,2 - 5	72,0 - 20	59,7	53,5 - 63,5
			Reverso	AAGACGACTCAGGGAGAGCA	60,1					
	Exón 39	247	Directo	AAGAGGCTGGAGTGCTCAA	60,1	95,0 - 5	53,0 - 5	72,0 - 20	58,7	54,9 - 64,9
			Reverso	CATCCATTGCCCTTCCCTAAA	59,9					
	Exón 40	235	Directo	CCTTGGGCTGGATGAGACTA	60,2	95,0 - 5	53,0 - 5	72,0 - 20	59,3	54,4 - 64,4
			Reverso	CCAGGAGTCACCAAGGACAT	60,0					
	Exón 41	225	Directo	CAGCATAGGCTGGAACTGC	60,9	95,0 - 5	54,0 - 5	72,0 - 20	56,7	54,0 - 64,0
			Reverso	GGCTGCTGATAGAACTGCT	59,6					
	Exón 42	231	Directo	AAGAAATGTGAAGGCATGATGA	59,6	95,0 - 5	53,0 - 5	72,0 - 20	59,5	54,2 - 64,2
			Reverso	CACATAGCAGGATCCAGGCTTA	60,8					
	Exón 43a	364	Directo	CCACCATTTTCTTGTCCAT	59,7	95,0 - 5	53,0 - 5	72,0 - 20	57,1	58,3 - 68,3
			Reverso	TCAACTCTAACAATGCCACAA	58,7					
	Exón 43b	384	Directo	AAACAACGACAAATCTTGAATCA	59,2	95,0 - 5	64,0 - 5	72,0 - 20	55,8	58,7 - 68,7
			Reverso	GGATTACAAGCATGAGCCACT	59,2					
	Exón 43c	399	Directo	TGGTACAGATGCTTTCGACAAT	59,6	95,0 - 5	60,0 - 5	72,0 - 20	59,2	59,0 - 69,0
			Reverso	CCTCCAGGTTCAAGTGATT	59,0					
FANCL	Exón 12	239	Directo	CCCTCCAAATATGTACAAG	62,8	95,0 - 5	53,0 - 5	72,0 - 20	54,5	55,6 - 64,6
	PHD Finger		Reverso	TCTATGTTGTGTTAGCGAAA	63,1					
	Exón 13		Directo	ATCATGCTGAGGCTTCTCT	63,9	95,0 - 5	53,0 - 5	72,0 - 20	56,5	50,0 - 60,0
	PHD Finger	158	Reverso	TGAAGCTAGGCACATTTTATGA	62,3					

VIII. ANEXOS
VIII.1. Anexo I: Información adicional

Gen	(continuación) Fragmento	pb	Sentido	Oligonucleótidos Secuencia 5'→3'	Tm (°C)	Condiciones de PCR (°C - segundos)				Condiciones de dHPLC (°C)	
						Desnaturalización	Anillamiento	Elongación	Tm	Gradiente	
E2F1	3' UTR-rs3213180	298	Directo	ACAGCCATGGGTGGTCAG	62,1	95,0 - 15	58,0 - 15	72,0 - 45	59,9	54,1-64,1	
			Reverso	GCCCAACGTCATCAAAATA	54,3						
	3' UTR-Site1-XbaI	555	Directo	ATACCGTCTAGATGTGACTTTGGGGACCTCAC	61,0	95,0 - 15	62,0 - 15	72,0 - 45	-	-	
			Reverso	ATACCGTCTAGAAAGAGTGTCTCTCAACGCTCT	59,1						
	3' UTR-rs3213180-Site2-XbaI	1036	Directo	GGCTAGTCTAGACAGGGCTTGGAGGGACCAAG	76,0	93,0 - 30	56,0 - 30	72,0 - 30	-	-	
			Reverso	GGCTAGTCTAGAAATATTTTAAACCAAGGAGGAGG	69,0						
	3' UTR-rs3213180-noSite2-short-XbaI	100	Directo	ATACCGTCTAGAGGGAGAGGTGAGTGGGGACCTT	79,0	95,0 - 15	64,0 - 15	72,0 - 45	-	-	
			Reverso	GGCTAGTCTAGACGCTTCAACACTGCAGGAG	75,4						
	3' UTR-rs3213180-noSite2-long-XbaI	187	Directo	ATACCGTCTAGAACAGCCATGGGTGGTCAG	75,3	95,0 - 15	64,0 - 15	72,0 - 45	-	-	
			Reverso	GGCTAGTCTAGACGCTTCAACACTGCAGGAG	75,4						
Firefly	3' UTR-rs3213180-C915G	-	Directo	AGGCTCTCCCGACGCCAGACGCTGTGTCTCTCTGCAAGTGT	92,8	95,0 - 50	60,0 - 50	68,0 - 300	-	-	
			Reverso	ACACTGCAGGAGGGACCAACAGCTGTGGCTGGGGGAGGCT							
	3' UTR-rs3213180-Site1-HindIII	-	Directo	TGGGGGGGCTCTAACTGGAGTGTGGGCCCTTTTGC	78,2	95,0 - 50	60,0 - 50	68,0 - 300	-	-	
			Reverso	GC44AAGGGCCGACACTCCAGTTAGAGCCGCCCA							
	3' UTR-rs3213180-Site2-HindIII	-	Directo	GCCCCACCTCCCAATCTGGAGTGTGATTGCTTCTTAAACAG	78,7	95,0 - 50	60,0 - 50	68,0 - 300	-	-	
			Reverso	CTGTTAGGAAGCAATCACTCCAGATTGGAGGGTGGGGC							
	E2F1-RT-PCR - #5	-	Directo	TCCAGAACCAACATCCAGTG	64,6	95,0 - 15	60,0 - 60	-	-	-	
			Reverso	CTGGGTCAACCCCTCAAG	63,2						
	Firefly-RT-PCR - #29	-	Directo	TGAGTACTTCGAAATGTCGGTTC	64,0	95,0 - 15	60,0 - 60	-	-	-	
			Reverso	GTATTCAGCCCATCGTTTCAT	62,6						
AmpR	AmpR-RT-PCR - #58	-	Directo	AGCTTCCCGGCAACAATTA	64,3	95,0 - 15	60,0 - 60	-	-	-	
			Reverso	GCTGCAAAATCTCTGGTTCTAA	63,2						
ACTB	ACTB-RT-PCR - #64	-	Directo	CCAACCGGAGAGAGATGA	65,1	95,0 - 15	60,0 - 60	-	-	-	
			Reverso	CCAGAGGGCTACAGGGGATAG	63,5						
			Reverso	ATCCGCTCAGGAGGTCTCTCAACACTCTGG							

Gen	SNP	Ensayo	Oligonucleótido directo 5'-3'	Oligonucleótido reverso 5'-3'	Oligonucleótido VIC	Oligonucleótido FAM	Sentido
ATM	rs1263936	rs1263936-V003	TGAGCGCAAGCCATCTCT	GTGGCATGCACTGTAGTCT	ACCTACTCCGAGGCT	CCTACTCCAGAGGCT	Reverso
	rs172896	rs172896-V005	CTTTAATATTTTGTCTTTTGCATATACCTCTAAATATCATC	TCTACCTATCTAATTTCTTGGGAAAAAAGATCAAAATAT	ACCTAGTGATTAATACTTTTAA	CCTAGTGATTAATACTTTTAA	Direto
	rs1801516	rs1801516-V004	GTCAAGCTGTACTTCCATCTGATTCA	CCCTGAAACATGTGTAGAAAAGCAGAT	TTTTTACTCCAGATACAAATG	TTTTTACTCCAAATACAAATG	Direto
	rs189037	rs189037-G004	CGTGGCTAAAGGAGAAAGAAAG	GGCTGGCAGGCGAGATC	CCTCCGCTGGCCAC	CTCCCGCGGCCAC	Reverso
	rs228589	rs228589-G003	CAGCATAGCCGGGTCCAA	GGCCGGCTTGTATTGGGTAA	CCCTCATCCCGCG	ACCTCTCTCCCGCG	Direto
BRG1	rs259164	rs259164-V001	AAAAAGAACTGAAACCCAGAGAGGTT	GAGCACTGATCAGGAGTTTGACTAT	CAACTCTGTGTAGGACAG	CAACTCTGTGTAGGACAG	Direto
	rs637064	rs637064-V002	GGCTGCCAGATAAATACCAT	AGGCAATCTTGTAAAGGCTCTTTATGG	ACGAAAAGGGTACTGTC	CGAAAAGGGTCTGCTGTC	Reverso
	rs1801514	rs1801514-S006	TCCTCTCATCTGCTTCCA	TCTTCCGCGAGCTGCATGAT	CCCTGTGTAACACC	AGACCTGTGATTAACAC	Direto
	rs1801376	rs1801376-S013	CTTCCAGTTTCACTCCATATGTG	TCGAAATATCCAAAACCCACACTCA	TGCACACAGCCAGT	TGCACACAGCCAG	Direto
	rs2070533	rs2070533-S008	CGTAAGGAAAAAGTAAAGGAAAAACGATT	CATTGATCATGCTATTGAAACAG	ATAGGCGGATGCTC	TCATAGGCTGATTGCTC	Reverso
E2F1	rs2071056	rs2071056-V021	AATCGTTCCGGGTAGCCATTATTAC	ACGCCCTGAGTTTCAAGTTCCA	ATCTGTATAATGAATCAC	TCTGTATAATGAATCAC	Reverso
	rs3213180	rs3213180-V020	GGATATGAGATGGGAGAGGTGAGT	GAGGGGCTTCAGACACT	CCGAGACCTCTGTGTC	CAGACGCTCTGTGTC	Direto
	rs2742976	rs2742976-V029	ACGAGAGACAGAGGGTGTGTT	GGAAATCCGGGCCCTTTTCTC	CACTGTGCATCCCGG	CACTGTGAATCCCGG	Reverso
	rs3213148	rs3213148-V028	GTGCAGCTCTGGGCTTAAT	GTCTCTCTCCACCTGTG	CACCCAGGCCCA	CACCCAGGCCCA	Reverso
	rs3218211	rs3218211-V026	AGCTGAAGAGCTGTGATGAACAC	CACTGTCTGAAAGCTCAGAGA	TTCGACCACTCATC	CTTGGACCACTCATC	Direto
E2F3	rs1042317	rs1042317-V033	GTTCCTCATGCGAAGAGATTTTTCAGTT	CTTAATGCAAAACAAATACAGTATGGACATCT	AATTAGCACTTGGCAGACT	CAATTAGCACTTGGCAGACT	Reverso
	rs1322842	rs1322842-V032	CTAGCGTTCTCTCCCATATCT	AGGAGAAAAAGCTTTTGTGGAGAAAT	TAGACCAAGTTAAGCCAG	CACCAAGTTAAGCCAG	Reverso
	rs4134957	rs4134957-V030	GAAAAAGGAAATTAACCTACCAAGAGTACTGAGA	GTCTGCGACTTGAATGATTTGAG	CCGCCAGTGAATG	CCGCCAGTGAATG	Reverso
	rs911361	rs911361-S015	TGGAGCCGGATCTTCAAGTC	TGGAAATGCTGCGTGACA	ACTGATTTAGGAAAGAG	ACTGATTTAGGAAAGAG	Direto
	rs3729639	rs3729639-V034	AGTCTCTGAGCACTGTACTGTGA	CCCATGGCTGCTGCTCATC	CCATACCACTACTG	CCATACCACTACTG	Reverso
E2F5	rs3808538	rs3808538-V038	CCCTTGAAGTATTGGGTATGAACAAATGT	GTAAGTCCCTTAATAGCTTGCTGT	CTTAAGACAGTAAATAG	CTTAAGACAGTAAATAG	Direto
	rs20552	rs20552-S012	ACCAAGAGATACGACAGGCACT	TTCTGGATCTGACGGTGAATG	ATGCCAATTTTGGAGG	TGCCAATTTTGGAGG	Direto
FANCA	rs2239359	C_2590906_1	-	-	-	-	-
	rs7195066	rs7195066-E001	GAGCGCAGGGACACAGGAA	GTCCCGAGAGGTGGATGTG	CACCTGTGCTGGG	CACCTGTGCTGGG	Reverso
	rs2905223	rs2905223-V093	CACCACAAAGAGAGACAGACATTC	CCCTAGCCAAAGCACCTAACTT	ATCAGTTATTAAATCTTTT	TCAGTTATTAAATCTTTT	Reverso
	rs5935816	rs5935816-V094	TGCTAAGAAATCAATTAACCATATCAAAAAAGATAACTG	AGCCAGAAACAAATATCTGAAGTCACTT	AATGTGTCAACAGCATAT	TGTGTCAACAGCATAT	Direto
	rs2272125	rs2272125-L1366L	CAACATGTGCTCTGCTGCTCAAAA	TGTTGAGAGTGAACATAGCTTTGAC	CCCTGGAATCTTATG	CCCTGGAATCTTATG	Direto
FANCD2	rs722509	C_626143_10	-	-	-	-	-
	rs9811771	rs9811771-V092	CTGTTGAGACTTCTAGTATCTCAT	TGGTTTCTGCGATTATTGGTGAGA	AGTATGAGAGCTGCTTGA	TATGAGAGCTGCTTGA	Direto
	rs33917318	rs33917318-V091	CAGCAGGATGAGACAGTATG	GGTGTCTCTACCCATGATCA	CTCTCTGATGATGATC	TTCTGATGAGCATG	Reverso
	rs3740615	C_3093636_10	-	-	-	-	-
	rs4447177	C_27920940_10	-	-	-	-	-
FANCF	rs3087374	C_15793548_10	-	-	-	-	-
	rs7168941	C_29177297_10	-	-	-	-	-
	rs3032440	rs3032440-V107	TCCAGAAATTGCCAGTGTGCTTAA	AGTGGCTTGGAGGACAGTAGATC	CACCCCTGGTTGAA	CACCCCTGGTTGAA	Direto
	rs2048718	rs2048718-V019	GGCTTCTCTCTCTCTCT	GTGGGTGAGGAAAGGTAACG	CCCCAGTCTGCACAC	CCCCAGTCTGCACAC	Reverso
	rs4986765	rs4986765-V016	CAGATTGAGCAGCACTCAACCTTG	AACCTTTTGTGCTTTTGGAAATCAAGC	CAGTGGAGTCTTG	CAGTGGAGTCTTG	Direto
FANCI	rs4983340	rs4983340-V018	CTTCGGTTCCCTCCAGACA	GCAGAAAGAACTCAAGCCCTTC	CAAGTCGGAGGGAAGT	AACTCGGAGGGAAGT	Direto
	rs4983351	rs4983351-V017	GTGCTGGGATTACAGGTGTGA	CTGAAAAAGGCTTAAAGAAACAACTAGA	CTTTGCCACCTCTT	TTGCCACCTCTT	Direto
	rs4983357	rs4983357-V015	CCCATTTACCTTTTGTAGAGAAGTCTTTCATTATTAG	ACCATCTGTTTCACTATTATTATTAACCT	CCATATTAAAGGAATTAATCTAT	CCATATTAAAGGAATTAATCTAT	Reverso

Anexo I. Descripción del diseño de las sondas utilizadas en TaqMan.

VIII. ANEXOS

VIII.1. Anexo I: Información adicional

(continuación)		Gen	SNP	Ensayo	Oligonucleótido directo 5'-3'	Oligonucleótido reverso 5'-3'	Oligonucleótido VIC	Oligonucleótido FAM	Sentido
FANCL	rs1011314	rs1011314-E005	TCAGGCTATGTGCTCTTTGG	GGGAATATAAATCTTATCAACAATCTAGGGGAAGG	CAATTTGTCAATATTGCG	CAATTTGTCAATATTGCG	CAATTTGTCAATATTGCG	CAATTTGTCAATATTGCG	Reverso
	rs3771224	rs3771224-E004	AGCAAAATAGCAGAAAAAGACAAACAGA	TCACATGGTAATTTGTTTTGAGAATTTTATTACTT	TCTTTTACTCTAACAGAAATAA	TCTTTTACTCTAACAGAAATAA	TCTTTTACTCTAACAGAAATAA	AATCTTTTACTCTAAATAGAATAA	Directo
	rs848289	rs848289-E006	CAGATGACAGGTGTTGAGGGCTAATA	GACTAGCCATTCATGAGCAATATAAATAATAA	CAATTTGTAACACTTTATTT	CAATTTGTAACACTTTATTT	CAATTTGTAACACTTTATTT	TTTGTAACACTTTATTT	Reverso
	rs226981	rs226981-V098	TCACCTAAGCTCTGAAAGGCAATAGA	CTTCCCTGTTTCCCTATGGA	CCTTCCCTGTTTCCCTATGGA	TCCACAGTAGACTTAA	TCCACAGTAGACTTAA	CCACAGTGCGCTTAA	Directo
FANCM	rs226984	rs226984-V097	GGTTTTGGCTTGAGCTGTTTTGG	ACTCAAACTGTTTCAACACTTACCA	TGTGTACAGACTTTA	TGTGTACAGACTTTA	TGTGTACAGACTTTA	TGTGTACAGACTTTA	Directo
	rs3736772	rs3736772-V099	CTTACAAGAAATACAGGTTCTTTTCTCT	CCACATTTAGCTGGGACACATATCTT	CTTGCGGAAGTCTA	CTTGCGGAAGTCTA	CTTGCGGAAGTCTA	CTTGCGGAAGTCTA	Directo
	rs8020533	rs8020533-V096	CTATTTTCTTTTATACCTTGGGCTTCTG	GGAGAATCAACTTCACTATTTTTCTGATCCT	CTGAGACTTTTGTGCTT	CTGAGACTTTTGTGCTT	CTGAGACTTTTGTGCTT	CTGAGACTTTTGTGCTT	Directo
	rs16940342	C_33566058_10	-	-	-	-	-	-	-
FANCN	rs249941	C_11716518_10	-	-	-	-	-	-	-
	rs447529	rs447529-V105	ACTTACAGCATCTCTCTATATATATCACTACACT	AAAGTTCCCTGCGCTCCGATTTT	TTGTTATATTCAATGATCTTC	TTGTTATATTCAATGATCTTC	TTGTTATATTCAATGATCTTC	TTGTTATATTCAATGATCTTC	Directo
	rs8058061	C_25742980_10	-	-	-	-	-	-	-
	rs7350	rs7350-EL70	GCCGGGAGGCGTTTGGT	TCCCAGGCGGAGAG	ACCGAAATCGGCGATC	ACCGAAATCGGCGATC	CCGAAATCGGCGATC	CCGAAATCGGCGATC	Reverso
H2AX	rs7759	rs7759-EL71	ATGTGGCCCCGCTGTAG	AGGCTCCAGGAGCACTTAA	TTTACTAATCAGAGCACTTAA	TTTACTAATCAGAGCACTTAA	TTTACTAATCAGAGCACTTAA	TTTACTAATCAGAGCACTTAA	Directo
	rs8551	rs8551-EL72	TGCAGAAAGCCATAGGTAAATATCA	TGGCATCTCTGGCGGCTTT	TTGCCCTCTGCTGACT	TTGCCCTCTGCTGACT	TTGCCCTCTGCTGACT	TTGCCCTCTGCTGACT	Reverso
	rs3730646	rs3730646-V009	GAATAGTCTATCTTTATGCTCTGCTTAAACA	AGTAAGAGGTATTTCCGTAAACCATTTTGAG	TTGGCTCACAATGA	TTGGCTCACAATGA	ATTGGCTCAATAATGA	ATTGGCTCAATAATGA	Reverso
	rs3730654	rs3730654-V010	GTAAATTTCCAGTATACCAATATATACTATAGTCTCA	AGGGCTCAATATCAAGTGTTCATGT	ATAGTCACTGTAAACAGCA	ATAGTCACTGTAAACAGCA	TCAGTGTAAACAGCA	TCAGTGTAAACAGCA	Reverso
MDM2	rs937282	rs937282-V006	GCTCGGGGCCAACAC	CACCACCTCCGGGATGATG	CCTCAGAGCCCGCC	CCTCAGAGCCCGCC	CCTCAGAGCCCGCC	CTCAGAGCCCGCC	Directo
	rs937283	rs937283-V007	CGAGCGCCCAAGTGC	GCGCGGCGAAGCA	AGAGTGGAAATGATCCCGAG	AGAGTGGAAATGATCCCGAG	AGTGGAAATGATCCCGAG	AGTGGAAATGATCCCGAG	Directo
	rs1052133	rs1052133-V014	CTCTACAGGTGCTGTTCAAT	ACCCCTTCTGCGCTTTGCT	CCAATCCCGCCATGC	CCAATCCCGCCATGC	CAATCCCGCCATGC	CAATCCCGCCATGC	Directo
	rs125701	rs125701-V012	GTCTCTGCTGGGACACAG	CAGACTCAAAATCTTGAAGAGCAA	CAATCCGCTGCAAGATG	CAATCCGCTGCAAGATG	TCAATCCGCTGCAAGATG	TCAATCCGCTGCAAGATG	Directo
OGG1	rs159153	rs159153-V011	CCCCTGTTGCCATGCA	AGGATTCAGCAAGCTGAAATGTCT	AAGGTGCTCAACACAG	AAGGTGCTCAACACAG	AAGGTGCTCAACACAG	AAGGTGCTCAACACAG	Reverso
	rs3219008	rs3219008-V013	GGCAGTGTATGTAATGGTTAAGA	GCTCCAAAAGTGTGCTGAATATACC	CCACCTGGCTAAC	CCACCTGGCTAAC	CCACCTGGCTAAC	CCACCTGGCTAAC	Reverso
	rs1573601	rs1573601-V039	CTGACATTTGCCCAAGAAAGATT	GCGAAGCACTGGCTACTCTC	AACAGCCAGTCTCTAC	AACAGCCAGTCTCTAC	AACAGCCATTCCTAC	AACAGCCATTCCTAC	Reverso
	rs198577	rs198577-V041	CTGCAGCTAGTGTCTCTCCAT	CGCTGAGGCCACAGAG	TCTCAGGCGAGAGCAG	TCTCAGGCGAGAGCAG	TCTCAGGCGAGAGCAG	TCTCAGGCGAGAGCAG	Reverso
RBL1	rs367438	rs367438-V042	GGTTTCTATGTTGGCCAGATTGG	CAGTTCTTTGGAGGCTGAGA	AAGCCAGGAATTCATA	AAGCCAGGAATTCATA	AGCCAGGAATTCATA	AGCCAGGAATTCATA	Reverso
	rs10748	rs10748-V043	GGGATCTCTGTGCAAACTAGA	TGGAGAAATCAAAAGCAGGTCACAG	ATTTTTCTCAGTTTCATC	ATTTTTCTCAGTTTCATC	TTTTTCTCAGTTTCATC	TTTTTCTCAGTTTCATC	Reverso
	rs1044377	rs1044377-V060	AATATTTGACATTCATACATTTGTTGGG	GCCCATCATGCTCCAGATTGG	ACATACGAAATACCTTT	ACATACGAAATACCTTT	CATACGAAATACCTTT	CATACGAAATACCTTT	Reverso
	rs1047972	rs1047972-V059	GGGCTCTGTGCTTCAAAATCTTC	CGGCTTGTGACTGGAGACA	CAGCGCATTCCTTT	CAGCGCATTCCTTT	AGCGGCTTCCTTT	AGCGGCTTCCTTT	Directo
STK6	rs2273535	rs2273535-V058	CTGGCCACTATTTACAGGTAAATGGA	TGGAGGTCCAAAAGTGTGTC	ACTCAGCAATTTCTT	ACTCAGCAATTTCTT	CTCAGCAATTTCTT	CTCAGCAATTTCTT	Reverso
	rs583609	rs583609-V102	GCTACTGGATATTTAGTCTCTCTAGGT	ACTAGTCTCTACTGAGGAAAATGAGTTCT	CAGTAATTCAAATTTCA	CAGTAATTCAGATTTCA	CAGTAATTCAGATTTCA	CAGTAATTCAGATTTCA	Reverso
	rs614665	rs614665-V101	CCCAAGGTCACAGAACTACTTAGGAT	CAGATCAAGACCTTGTCTCCAAA	CCGCTCTTGGCCAC	CCGCTCTTGGCCAC	CCGCTCTTGGCCAC	CCGCTCTTGGCCAC	Directo
	rs9436223	rs9436223-V100	CGCCGCTGTCTTAAACC	TGACAGACTCGATTCCGATTTT	AATCTACAGGACATCTT	AATCTACAGGACATCTT	AAATCTACAGGACATCTT	AAATCTACAGGACATCTT	Reverso
USP1	rs10735810	rs10735810-V054	GGGTACGCGAGGGAAGTG	TGGCCTGCTTGTGCTTCTT	TTGCTCTGCTGCTGCT	TTGCTCTGCTGCTGCT	TTGCTCTGCTGCTGCT	TTGCTCTGCTGCTGCT	Directo
	rs4516035	rs4516035-V057	GCAATGATGACCTTCTTTAGC	GCCATGTTTCTCTCTCTGTAAG	AATAGCAATATCTTCCC	AATAGCAATATCTTCCC	AGCAATGCTTCCC	AGCAATGCTTCCC	Reverso
	rs731236	rs731236-V055	CCGTGCCACAGATCGT	GTAGTCTGCGAGTGTGTTGGA	CGCTGATCGAGGCCA	CGCTGATCGAGGCCA	CGCTGATCGAGGCCA	CGCTGATCGAGGCCA	Directo
	rs739837	rs739837-V056	ATTCTCTCTGTTCTCTCAAC	GCGCCTTGCCTAAGTG	CCGCTGCTAAGTG	CCGCTGCTAAGTG	CGCTGCTAAGTG	CGCTGCTAAGTG	Reverso

Anexo I.7. Descripción de los oligonucleótidos utilizados en Amplifluor.

Gen	SNP	Ensayo	Oligonucleótido Directo	Oligonucleótido Reverso
FANCC	rs879559	rs879559-K-405	FAM - GAAGGTGACCAAGTTCATGCTACTCCTAGTGGACTCGGACC JOE - GAAGGTCGGAGTCAACGGATTACACTCCTAGTGGACTCGAACA	CCTCACGTCACTACACTAGAGT
FANCE	rs2395626	rs879559-K-405	FAM - GAAGGTCGGAGTCAACGGATTGTTCTTTACCAAATTGTGAGTTC JOE - GAAGGTGACCAAGTTCATGCTGTGTTCTTTACCAAATTGTGAGTTA	GACACAGAGGAAGGAAACACTA
FANCI	rs1557720	rs1557720-W-501	FAM - GAAGGTGACCAAGTTCATGCTGCCTGGTGTGAAGAATACAGAGT JOE - GAAGGTCGGAGTCAACGGATTGCTGCTGTGAAGAATACAGAGA	CTCCCTCACCCAGGTTTACTT

Anexo I.8. Descripción de los oligonucleótidos utilizados en iPLEX.

Gen	SNP	Oligonucleótido Directo	Oligonucleótido Reverso	Oligonucleótido de Extensión
ATR	rs12635931	ACGTTGGATGGAATTCAGAATAGCCAGCG	ACGTTGGATGACTCCTCTGTTTCAGCTACC	TCGGTATCCACTAAAAGACT
	rs6782400	ACGTTGGATGGACTGAAAGCACATGTGAAA	ACGTTGGATGATATCTCAGTACATGGCACC	CTAGATTTCCAGTTTGGAGAA
	rs7636909	ACGTTGGATGCCCAACCACTACTATTCC	ACGTTGGATGAGATTCAAGGTGATTCTCTC	TGTTTCCCAATTGCTAT
FANCC	rs1045276	ACGTTGGATGCAATAGCTGCTAGGAGTACC	ACGTTGGATGCGAGGTGAGCTGAGACTTTC	ACGCACCCACTCTAACT
	rs4647554	ACGTTGGATGCTCCGTTTTGAGGTATGAGC	ACGTTGGATGGGATCTATTAGCATTGCCAC	CAAAGTTTTGCTGGATTAGAAG
	rs554879	ACGTTGGATGGGATGCTAAAAATAGTTGCTG	ACGTTGGATGCTTCCCTAAATGCAGCATTTC	CGGTTTAAGGATTGTATTAGTGTGA
FANCE	rs760782	ACGTTGGATGCTAGAACCTGTATGTGGCTG	ACGTTGGATGCCATGGCTGTGCCTCCTCA	CCCTCGCCCCGCCAGAGAGGAAAAACA
FANCG	rs1053318	ACGTTGGATGGTTCAGCTACAGTCTGACTC	ACGTTGGATGGCTACCAATGAAAATCGC	TCTGACTCTGGACAGGG
	rs2258240	ACGTTGGATGAAGGCCCACTAGAAAAGGAG	ACGTTGGATGTCAGGTACACACATACGTGC	GGGGAAACTGGGGATGAGACTTA
	rs568300	ACGTTGGATGCTCTCCATCTCAATGCTCAG	ACGTTGGATGATGCTTCTTTCATGCCACC	CCCCAATGCTCAGCAGGTAGGT
FANCI	rs7220740	ACGTTGGATGCAGACTTGAAAGTGGTTCAG	ACGTTGGATGAAATGGCCATAACCCCAAG	GGGAAAGGGCATGAGGCAAC
	rs8080599	ACGTTGGATGGCCTCCATTCTACTGTAAGG	ACGTTGGATGGTGACTGTGTGACTTTAGGG	ACAATGCTGAAAGACATTTATTAA
	rs9897928	ACGTTGGATGCATACAAATTTAGTGCTGAAC	ACGTTGGATGTCAGTGAATGTCAGATGAGG	AACTACAGCTAGAGAGTGA

VIII. ANEXOS

VIII.1. Anexo I: Información adicional

Anexo I.9. Descripción de los oligonucleótidos utilizados en *Illumina GoldenGate*.

Gen	SNP	Ensayo	Oligonucleótido 1	Oligonucleótido 2	Oligonucleótido 3	Sentido
ATM	rs170548	rs170548-37_B_F_794415	CACGCTACTAGATTAGCAGAAACCAAAATCT	CAGCTACTAGATTAGCAGAAACCAAAATCG	TCTATTTTAAATGCTGTTTGGAGAAATCGGCCCTCTAGACTCGT	Reverso
	rs1300058	rs1300058-120_I_F_794434	GCTGCTTGCTCACATATGGCAGAA	GCTGCTTGCTCACATATGGCAGAA	AGAATTTTCTGATAGGAAATCTCTAGACTGTTTAAACCGTCTGGCAGAC	Directo
	rs1801516	rs1801516-116_I_F_794455	GCATCTTGATCATGATATTTTACTCCAAA	GCATCTTGATCATGATATTTTACTCCAAA	TACAAATGAATCAGGAGAAATGCGACATTCGGCCGACAGATGTCT	Directo
	rs227060	rs227060-113_I_F_794551	CACAAGGCAATGAGAAAGCAAGCAA	CAGAAGGCAATGAGAAAGCAAGCAG	AGAAAGCTTTTGATTTTAAAGGATGAGATCGGTTTCAGTCTCCGG	Directo
	rs227064	rs227064-37_B_F_794552	TTTCATACATGCGCAGGCAAAAT	TTTCATACATGCGCAGGCAAAATC	GTAGCATGGATTTCACAAAAGTTGTGGCGTCAAAAGCTAGGAGATCG	Reverso
	rs228590	rs228590-37_B_F_794560	GATGGCTGCTGATCTCTCTCTTTT	GATGGCTGCTGATCTCTCTCTTTT	TATTCATGCTGGTCTTTCTTCTGACGCAAAATCTGTCTCCGGTA	Reverso
	rs228599	rs228599-120_B_F_794561	AAGTTCGTGTTATGAGAGACTTTGACAT	GAGAGTTCGTGTTATGAGAGACTTTGACAC	TGACTTAAAAATCATCTGTTTAAATGTCAGCGCTCTGTATCTACTAGGCTCT	Reverso
	rs4585	rs4585-103_B_F_794830	AAGCAAGAGAGAAAAACTTTTGACAT	AAGCAAGAGAGAAAAACTTTTGACAC	TAAAAAAGAAAGCCAGATATTTGGTTCATGACACGACGACGCAACTG	Reverso
	rs625120	rs625120-37_B_F_794906	TGAGTGTCCACATATAAGATGAACTTCAT	TGAGTGTCCACATATAAGATGAACTTCAC	TATACATTAATTCAGAAATATCTAGGTTTACACGTAAACCATCTGGATACACAG	Reverso
	rs660429	rs660429-117_B_F_794920	GAATGTGACTGATTTTATTTCTGTGACTAA	GAATGTGACTGATTTATTTCTGTGACTAC	TAAATTCATTTCTTCTGACTGACTATATCTGTGTATGTCACGAAAGCGGA	Directo
BRCA1	rs664143	rs664143-113_B_F_794921	GGCTCTTCACATCAGTGGTCTCTGAT	GGCTCTTCACATCAGTGGTCTCTGAC	AGATACACATCTCATCTGTGAAGAAATCAATGACCGTCTATAGATGCGCA	Reverso
	rs16941	rs16941-117_B_F_794413	GTTCATTAAATATGCTTGAGCTGGCTT	GTTCATTAAATATGCTTGAGCTGGCTTC	TTTAAAAACAATTTCTCTAATGTATCTTTCACCCGGCTCGATGAA	Reverso
	rs1799949	rs1799949-117_B_F_794430	GTGTTAAATTCAGTCTGGGAAAGATCA	GTGTTAAATTCAGTCTGGGAAAGATTCG	TGTCATCTCTTTTACTGTCTTTTCTTACTCATCTCATCTAGACGAC	Directo
	rs1799966	rs1799966-117_I_F_794431	GTGAAAAGTCGAGAATCTGCCAGA	GTGAAAAGTCGAGAATCTGCCAGG	TCCAGTCTGCTCATCTACTATAGGAAATCTGATCCACGCGG	Directo
	rs799906	rs799906-117_I_F_794994	GCGCAGGCGCACTGAACCTTACAA	GCGCAGGCGCACTGAACCTTACAC	ACTATGTCGCCGACCTCTGTTCACGCGCATCTCTTAGAGAA	Directo
	rs799923	rs799923-96_B_F_794995	GGAAAAAGCAGACGAACTGGCCAT	GGAAAAAGCAGACAGAACTGGCCAAC	ATTGCTGACTGTTCTTTACCATCTCTCAATCTTATTCGGCAGGTGA	Reverso
	rs817619	rs817619-117_I_F_795002	GAATGAATAGAGAGACTTTTGCAATTTTA	GAATGAATAGAGAGACTTTTGCAATTTTC	ACTTATAAATTTTCCGCTCTGTCGACGCGCTCTAATGTAACCTCTG	Directo
	rs8176265	rs8176265-117_B_F_795003	GTGAAATATTCTCTCTCTACACTTTAT	GTGAAATATTCTCTCTCTACACTTATC	TATTTACTCTCACTCTCTCAGTATGCGGCTGGTTTCTACCCACAG	Reverso
	rs8176318	rs8176318-119_I_F_795004	GCCATTGAAGGGTCTGACTCTGCGA	GCCATTGAAGGGTCTGACTCTGCCC	TTGTGAACACAGGGTTTATGAGATGACCATCATCGGAGCGTTGAT	Directo
	rs15969	rs15969-37_B_F_794408	GGCTGAGGTAAATTTGAAATCAATTTTCTT	GGCTGAGGTAAATTTGAAATCAATTTTCTG	TCAATTTGCAACATTAAGTATTAATGACCACTGCGGTTCTACATG	Reverso
BRCA2	rs1801406	rs1801406-119_B_F_794453	TGTACTCTCTCTGCAATATGAGCTTGCT	TGTACTCTCTCTGCAATATGAGCTTGCC	TTCTAAACTGAGTAAATCAACTGCGGTGTGTGTGATAGATGCGCA	Reverso
	rs1801439	rs1801439-119_B_F_794454	GCACGTGTTCTCTCATTTATGCGCTTCT	GCACGTGTTCTCTCATTTATGCGCTTCC	ATTTTGTGAGCTAGAAATACATGATCTGCTGCTGTGTGATAGATGCGCA	Reverso
	rs206079	rs206079-116_I_F_794497	GAATTAATAGGTTCCAGTTACATCTAGAA	GAATTAATAGGTTCCAGTTACATCTAGAG	ATGAACAATATCTAAAGCTTAGCTTGCGAATCTCCGTCAGGAAAGCGT	Directo
	rs206116	rs206116-101_I_F_794499	AGCCCAACCTTTTCTTAATCCAAA	AGCCCAACCTTTTCTTAATCCAAAG	TGCTCTCTGTTTACTGCTCAAGCTCTGACGATCCGAGATCGAGGTCC	Directo
	rs206146	rs206146-37_I_F_794500	GTGGGGCAAGGAAATGAGAGGA	GTGGGGCAAGGAAATGAGAGG	AAGGAGGAATGATGAAATATAGGAGCTTAAAGCGGCGGCTCACTTC	Directo
	rs3092909	rs3092909-119_B_F_794669	GGCTACTCGAAGTCCCGCATCT	GGCTACTCGAAGTCCCGCATCC	TTTCTTTTGTCCAGCTCATGAAGGGCTGCGGATGTGTGACTGTT	Reverso
	rs3764791	rs3764791-119_B_F_794763	ACCTCCCATCTCTACTCGGTTAT	ACCTCCCATCTCTACTCGGTTAG	TAGTTGTAAACATCACCCACACGCGTCTGAACCTAGCTGTATCTCC	Reverso
	rs9534262	rs9534262-219_I_F_795048	TCACAGAGAGCCCTGAAACAAAT4A4A	TCACAGAGAGCCCTGAAACAAAT4A4AG	AAACACCATAGGATGATGATCTGAT4A4ACATGAT4A4TACTTGGGCTCAGA	Directo
	rs1801376	rs1801376-116_I_F_794458	CTCCATATGTGGAAGAGACTGCA	CTCCATATGTGGAAGAGACTGCAAC	CAGCCAGTATGTGATGATGTGATGCTGTAGAGCGTGAATCCCG	Directo
	rs2305653	rs2305653-340_B_F_794578	GGTTGTTTGAAGCAGGAAGGCAAT	GGTTGTTTGAAGCAGGAAGGCAAC	TGTTTTTTAATAACAAAAGCTTCCGACTCCGACTAGCAGATGCTGTGAC	Reverso
BUB1B	rs3214017	rs3214017-116_B_F_794700	GTGGGTGATATAGTGGAGTACACAAAT	GTGGGTGATATAGTGGAGTACACACAG	CCTAAGTCTTTTGTAGCTGTAGAGACGCTGATGTGCTGTGTG	Reverso
	rs3759839	rs3759839-117_B_F_794759	GGAAAGACTGAGATCAAAAGACAAACA	GGAAAGACTGAGATCAAAAGACAAAC	TACGTTTCCAAAGTTTGTGTGTGTAGGATGACGAGATGTGCCCTGTGTGA	Directo
	rs9424437	rs9424437-119_I_F_794869	GTCACTCTAGCTATCATTTCTTGAAA	GTCACTCTAGCTATCATTTCTTGGAAT	AGACAATATCACTGAAGCTTCATTGAGGACAGAAACGTAAATCTCCGC	Directo
	rs3213150	rs3213150-114_B_F_794695	GGGGGTGGAGGGGGGTGTACT	GGGGGTGGAGGGGGGTGTACCC	CTGTGCTCTGTGATCCAGGCGACACAAATGTTGTGTCCGAG	Reverso
	rs3213183	rs3213183-349_I_F_794696	GCCATCGTGTCTCTGCTGCTGGTA	GCCATCGTGTCTCTGCTGCTGGTG	TTTCTCTCTGCTGTTAACTCTCTAGACATGGGCTGCTCCGTAACAT	Directo
	rs3238489	rs3238489-100_I_F_794583	GACACGTGCTGGTCCCAAGTGCTC	GACACGTGCTGGTCCCAAGTGCTG	ATTGGAAGAAAGACAGATGCTGAGGATGATGCTGCTCCAC	Directo
	rs4712494	rs4712494-111_B_F_794583	ATGCAAGGCCCACTTAGCTC	ATGCAAGGCCCACTTAGCTC	GTAAATCAAGCGGGTTTATCAAGACGCGGATCTATTACTCTGCG	Reverso
	rs1057042	rs1057042-219_I_F_794330	GTCTCCCTCTCGGGGGGCTCA	GTCTCCCTCTCGGGGGGCTCG	GCGATCTCTCTGCTGATGCTGATAGCGGGGTCTGTACCACTG	Directo
	rs1800359	rs1800359-116_I_F_794435	GTGTCAATTTCTCATGTCCCCACA	GTGTCAATTTCTCATGTCCCCACG	CATCTGTGACTGTACCTCTGCAAGCTCGGTTTCCGTGGACC	Directo
	rs2238526	rs2238526-119_I_F_794537	GTTGCCGCGACGACAGCAGCA	GTTGCCGCGACGACAGCAGCG	GTTTCTTTTGTGGCCCTATACAGTTGCGTCTCACTAGTGTG	Directo
FANCA	rs2239359	rs2239359-119_I_F_794538	GGGCTCATTTATTAAGTTTGTGTGTC	TGAGGAGGAGGAGCGTACTTTGCC	ATGTGCACCTCAGGATAGATGTGTGTCATACAGACCGGAGAC	Reverso
	rs23768794	rs23768794-116_I_F_794587	GGGCTCATTTATTAAGTTTGTGTGTC	TGAGGAGGAGGAGCGTACTTTGCC	TGTGAGCTCAGGCTTCTTATATACACAGGACCGGACGCTTTC	Directo

(continuación)		Ensayo		Oligonucleótido 1	Oligonucleótido 2	Oligonucleótido 3	Sentido
FAMCA	rs3785275	rs3785275-116_B_F_794778	CTGTGCGAGCTCAAGCCGG	CTGTGCGAGCTCAAGCCGG	CTGTGCGAGCTCAAGCCGG	GCTTTGGGTACAGGGTCTCACTGTTTCAAGTGCATCAGCG	Reverso
	rs7190823	rs7190823-119_B_F_794955	GCTTCTCTGCGAGTACATCAACCGT	GCTTCTCTGCGAGTACATCAACCGT	GCTTCTCTGCGAGTACATCAACCGC	ACCTGTCAAAATAGAAATGATGATTACTAGTCAATGAGCGGATTTCTA	Reverso
	rs7195066	rs7195066-119_I_F_794956	ATGTGGGTCTCTCTGCGACCTGA	ATGTGGGTCTCTCTGCGACCTGA	ATGTGGGTCTCTCTGCGACCTGG	CTCTTTGACAGCGCTCTCTGACCCGGTATCTTTGAGCGATATACTT	Directo
	rs1144943	rs1144943-37_I_F_794347	GTGGCTTTGTGCTTTTGAAGAAATTA	GTGGCTTTGTGCTTTTGAAGAAATTA	GTGGCTTTGTGCTTTTGAAGAAATTC	CTTGACAAAATATTAGAGCAGAACTGACTGTACCGCTAATCGCG	Reverso
	rs1470383	rs1470383-119_I_F_794386	TCCTTGCTCTTTTGGATTGGATCTA	TCCTTGCTCTTTTGGATTGGATCTA	TCCTTGCTCTTTTGGATTGGATCTG	CTGTGATCACAACTGATATCTTGGATTACCAAGGGGCGTCAATTGG	Directo
MDM2	rs1201644	rs1201644-94_I_F_794354	GCAGAAAGATGCAATGATATATATATGGA	GCAGAAAGATGCAATGATATATATGGA	CAGAAAGATGCAATGATATATATATGGA	AAAGTCAATAAGTGTGTTTACAACTAGAGTCAACGATTTTGGCCGGAC	Reverso
	rs3730590	rs3730590-107_I_F_794730	TTTTATCGAGATCACAAGTTCTGCACA	TTTTATCGAGATCACAAGTTCTGCACA	TTTTATCGAGATCACAAGTTCTGCACG	GAGTCTTAATCATGTTTGTGTTTCAAGCGGTGATAGGGGCGAGCTCAAT	Directo
	rs1695144	rs1695144-116_I_F_794414	GGCCCGAGGAAACCAAGGGA	GGCCCGAGGAAACCAAGGGA	GGCCCGAGGAAACCAAGGGG	TGATATCATTTTTTGGTAAATATGCTAGAACGCTGATGCTTGGATGGAGCC	Directo
	rs3730646	rs3730646-107_B_F_794731	GGCTAACCATTTTGAGATTGCTCAT	GGCTAACCATTTTGAGATTGCTCAT	GGCTAACCATTTTGAGATTGCTCAC	AATGATGTTTAAAGACAGGAGCAATCTGCTGATGCTTGGATGGAGCC	Reverso
	rs1894255	rs1894255-119_I_F_794469	TTAGGCTTTTCAATTAATCAGTATCAT	TTAGGCTTTTCAATTAATCAGTATCAT	TTAGGCTTTTCAATTAATCAGTATCATG	TAAATGTTTGTATAATTTCTTACCAATGTTGTCACCGCTTTTGGACGA	Directo
RB1	rs1951775	rs1951775-92_B_F_794475	GTCATGTTTAACTTAATTCACAAATCTT	GTCATGTTTAACTTAATTCACAAATCTT	GTCATGTTTAACTTAATTCACAAATCTG	GTATGAGCAATTAATAGATAATTGTAACTTGGCTCCGACCTAAG	Reverso
	rs198575	rs198575-87_B_F_794481	AAACCAAGCTATCTAATCCCAAGTCT	AAACCAAGCTATCTAATCCCAAGTCT	AAACCAAGCTATCTAATCCCAAGTCC	GGCTCTTAATCACTTAATCAATCTCAGATACCGAGAGACTTGGCC	Reverso
	rs198580	rs198580-87_I_F_794482	TGCTACTTAACAGCAATTAATAGAGCA	TGCTACTTAACAGCAATTAATAGAGCA	TGCTACTTAACAGCAATTAATAGAGCG	TTTCATGATTTGAAAAAATCTACTCAACCATGTGCGAGCTTCTGG	Directo
	rs198590	rs198590-110_I_F_794483	GACTAAATGTCTACTCTTGGCAATGCA	GACTAAATGTCTACTCTTGGCAATGCA	GACTAAATGTCTACTCTTGGCAATGGG	AATGTATCTGAAACCCCTTATAGATTGGCGCTGCTTGGACGCAAAAGTTA	Directo
	rs2804094	rs2804094-110_B_F_794644	GAGCAAACTCCCGACTATGCTT	GAGCAAACTCCCGACTATGCTT	GAGCAAACTCCCGACTATGCTCG	AGGTTCAAAACGCTGCTGATTTGGGACATGCGCTCTCTGTAAAT	Reverso
5TK6	rs399413	rs399413-110_B_F_794795	GCAGGTTATGGCAATGAATGAATGT	GCAGGTTATGGCAATGAATGAATGT	GCAGGTTATGGCAATGAATGAATGTC	AACAAAGTACAAAAGACACAGTTGCGCGGTATTGTGGGCGACAGGAT	Reverso
	rs4151441	rs4151441-110_I_F_794803	AGAGGGAAGCAAAAAGTTAATTGCTA	AGAGGGAAGCAAAAAGTTAATTGCTA	AGAGGGAAGCAAAAAGTTAATTGCTG	TTAAAAGTCAAGCAGTTACTTCTAGTCTAAGTTCCGGTTCCGACGTG	Directo
	rs4151474	rs4151474-110_I_F_794804	GGCACTGCTGCGAGGCGCAA	GGCACTGCTGCGAGGCGCAA	GGCACTGCTGCGAGGCGCAAT	CCCAAGATTACTCAATCTGATGTTTCCGAAAGCCCAACGCTGATTCTGGGC	Directo
	rs520342	rs520342-110_B_F_794837	TAAGCTCTGCGCATAAAGGAAAT	TAAGCTCTGCGCATAAAGGAAAT	TAAGCTCTGCGCATAAAGGAAAC	AATTTCTGGATTCTCTCAATGAAGAGCACAACGCTGATTCTGGGC	Reverso
	rs9568029	rs9568029-119_B_F_795049	GAGGACGCTCTGCTGCGCAAGT	GAGGACGCTCTGCTGCGCAAGC	GAGGACGCTCTGCTGCGCAAGC	TTTATGCTGGGTTCTGTAGATATTAGTGCCAAAATGCCCCGTGAG	Reverso
5TK6	rs9568036	rs9568036-119_B_F_795050	GCCATCAATTTTGAATTTCAAAAGAAAT	GCCATCAATTTTGAATTTCAAAAGAAAT	GCCATCAATTTTGAATTTCAAAAGAAAT	AAAAAGCTTAAAAATTTTGTATTAAGGATCGGAATGACAACTCACCAAGT	Reverso
	rs1044377	rs1044377-114_B_F_794695	CATTGGTTGGGGGAAAGAAAGTAT	CATTGGTTGGGGGAAAGAAAGTAT	CATTGGTTGGGGGAAAGAAAGTAG	TCTGGTATGTTTCATGATCTGTAGCCCTACCGACGCGGTGTA	Reverso
	rs1468056	rs1468056-114_I_F_794384	GTGGAGCAAAATGCGCTGAGCTTCTC	GTGGAGCAAAATGCGCTGAGCTTCTC	GTGGAGCAAAATGCGCTGAGCTTCTG	CAGAAAGATGTTGTAATTAATGTTGTACGCAAAAGAGGCGGTCTGATG	Directo
	rs6014711	rs6014711-114_B_F_794897	GACCCAGGAGGAGACACACAT	GACCCAGGAGGAGACACACAT	GACCCAGGAGGAGACACACAC	AATTTGGTCTCTTTGATTGAGTACCAAGTCTTACGCTCAGCGG	Reverso
	rs9111160	rs9111160-114_B_F_795030	GAGTCAACCTGCGAGGTAGCTGCTCG	GAGTCAACCTGCGAGGTAGCTGCTCG	GAGTCAACCTGCGAGGTAGCTGCTCG	AATTTCTGCTACACCTCAGAAATATCTGCTGACAAAGGTTTCTAGGG	Reverso

Anexo I.10. Descripción de los genes de reparación del ADN seleccionados para el estudio y las referencias bibliográficas que determinaron su elección.

Gen	Descripción	Referencias bibliográficas
ATM	Ataxia telangiectasia mutated	(Ahmed and Rahman, 2006; Angele <i>et al.</i> , 2003; Chenevix-Trench <i>et al.</i> , 2002; Koren <i>et al.</i> , 2006; Maillet <i>et al.</i> , 2002; Tamimi <i>et al.</i> , 2004; Taniguchi <i>et al.</i> , 2002b; Thorstenson <i>et al.</i> , 2003)
ATR	Ataxia telangiectasia and Rad3 related protein	(Andreassen <i>et al.</i> , 2004; Gatei <i>et al.</i> , 2001; Yang <i>et al.</i> , 2003)
BRCA1	Breast cancer type 1 susceptibility protein	(Cox <i>et al.</i> , 2005; Wang <i>et al.</i> , 2000b)
FANCA	Fanconi anemia group A protein	(Folias <i>et al.</i> , 2002; Thompson <i>et al.</i> , 2005)
FANCB	Fanconi anemia group B protein	(Fei <i>et al.</i> , 2005; Meetei <i>et al.</i> , 2004a; Rahman and Ashworth, 2004)
FANCC	Fanconi anemia group C protein	(Kitao <i>et al.</i> , 2006)
FANCD1/BRCA2	Fanconi anemia group D1 protein/Breast cancer type 2 susceptibility protein	(Howlett <i>et al.</i> , 2002; Powell <i>et al.</i> , 2002; Stewart and Elledge, 2002)
FANCD2	Fanconi anemia group D2 protein	(Houghtaling <i>et al.</i> , 2003; Hussain <i>et al.</i> , 2004; Ohashi <i>et al.</i> , 2005; Wang <i>et al.</i> , 2004b)
FANCE	Fanconi anemia group E protein	(Gordon <i>et al.</i> , 2005; Leveille <i>et al.</i> , 2006)
FANCF	Fanconi anemia group F protein	(Wang <i>et al.</i> , 2006)
FANCG	Fanconi anemia group G protein	(Hussain <i>et al.</i> , 2003)
FANCI	Fanconi anemia group I protein	(Sims <i>et al.</i> , 2007)
FANCI/BRIP1	Fanconi anemia group J protein/BRCA1 interacting protein C-terminal helicase 1	(Cantor <i>et al.</i> , 2001; Levitus <i>et al.</i> , 2005; Litman <i>et al.</i> , 2005; Seal <i>et al.</i> , 2006)
FANCL	Fanconi anemia group L protein, Ubiquitin ligase protein PHF9	(Marek and Bale, 2006; Meetei <i>et al.</i> , 2004b)
FANCM	Fanconi anemia group M protein	(Meetei <i>et al.</i> , 2005)
FANCN/PALB2	Fanconi anemia group N protein/partner and localizer of BRCA2	(Rahman <i>et al.</i> , 2007; Tischkowitz <i>et al.</i> , 2007)
H2AX	Histone H2A.x	(Bogliolo <i>et al.</i> , 2007; Fernandez-Capetillo <i>et al.</i> , 2004; Lyakhovich and Surrallés, 2007; Monteiro <i>et al.</i> , 2003)
OGG1	N-glycosylase/DNA lyase	(Rossner <i>et al.</i> , 2006; Zhang <i>et al.</i> , 2006)
USP1	Ubiquitin carboxyl-terminal hydrolase 1	(Nijman <i>et al.</i> , 2005)

Anexo I.11. Descripción de los SNPs seleccionados en los genes de reparación del ADN.

Gen	SNP	Cromosoma - - Posición	Posición génica	Localización génica	Cambio nucleotídico	MAF caucásico HapMap
ATM	rs228589	11-107598418	-351	5' corriente arriba	A>T	0,48
	rs189037	11-107599043	274	5'UTR	A>G	0,49
	rs625120	11-107600430	1661	Intrón 1	C>T	0,40
	rs228590	11-107601351	2582	Intrón 1	C>T	0,40
	rs228599	11-107612870	14101	Intrón 5	G>A	0,48
	rs599164	11-107625649	26880	Intrón 9	G>A	0,48
	rs637064	11-107646119	47350	Intrón 18	T>C	0,40
	rs1800058	11-107665560	66791	Exón 29 (Phe1420Leu)	C>T	0,06
	rs647681	11-107666462	67693	Intrón 29	G>A	0,38
	rs1801516	11-107680672	81903	Exón 37 (Asp1853Asn)	G>A	0,18
	rs660429	11-107686721	87952	Intrón 39	A>C	0,48
	rs227060	11-107710091	111322	Intrón 54	C>T	0,22
	rs227064	11-107712603	113834	Intrón 56	A>G	0,47
	rs172896	11-107722259	123490	Intrón 58	T>G	0,48
	rs664143	11-107730871	132102	Intrón 61	T>C	0,48
	rs170548	11-107740046	141277	Intrón 61	T>C	0,22
	rs4585	11-107744838	146069	3'UTR	G>T	0,47
ATR	rs6782400	3-143778058	2283	Intrón 1	A>C	0,40
	rs7636909	3-143743252	37089	Intrón 17	A>G	0,46
	rs12635931	3-143647433	132908	3' corriente abajo	T>C	0,11
BRCA1	rs799906	17-38531642	-648	5' corriente arriba	T>C	0,36
	rs799923	17-38505457	25537	Intrón 6	G>A	0,31
	rs1799949	17-38498992	32002	Exón 10 (Ser694Ser)	C>T	0,35
	rs16941	17-38497961	33033	Exón 10 (Gly1038Glu)	A>G	0,36
	rs8176199	17-38484050	46944	Intrón 12	A>C	0,26
	rs1799966	17-38476620	54374	Exón 15 (Gly1613Ser)	A>G	0,35
	rs8176265	17-38467522	63472	Intrón 18	G>A	0,35
	rs8176318	17-38450800	80194	3'UTR	G>T	0,35
FANCA	rs2376879	16-88412323	-1757	5' corriente arriba	C/G	-
	rs7190823	16-88393544	17022	Exón 9 (Ala266Thr)	T>C	0,44
	rs1057042	16-88385123	25443	Intrón 14	A>G	0,39
	rs2239359	16-88376981	33585	Exón 16 (Ser501Gly)	C>T	0,38
	rs3785275	16-88369530	41036	Intrón 20	G>C	0,39
	rs7195066	16-88363824	46742	Exón 26 (Asp809Gly)	C>T	0,23
	rs2238526	16-88354224	56342	Intrón 29	A>G	0,38
	rs1800359	16-88332762	77804	Intrón 42	C>T	0,33
FANCB	rs11795930	X-14798736	2369	Intrón 1	C>T	0,11
	rs5935816	X-14793869	7236	Intrón 1	A>G	0,39
	rs2905223	X-14778727	22378	Intrón 5	G>A	0,31
FANCC	rs554879	9-97107630	12175	Intrón 1	G>C	0,49
	rs4647554	9-96902522	217283	3'UTR	T>C	0,40
	rs1045276	9-96888505	231300	3' corriente abajo	C>T	0,11
FANCD1/ BRCA2	rs206116	13-31786483	-1134	5' corriente arriba	C>T	0,48
	rs3092989	13-31787363	-254	5' corriente arriba	G>A	0,10
	rs1801439	13-31804980	17363	Exón 10 (Ser455Ser)	A>G	0,03
	rs1801406	13-31809888	22271	Exón 11 (Lys1132Lys)	A>G	0,32
	rs206079	13-31818618	31001	Intrón 12	A>G	0,47
	rs9534262	13-31834646	47029	Intrón 16	T>C	0,43
	rs3764791	13-31849111	61494	Intrón 21	A>C	0,23
	rs206146	13-31854757	67140	Intrón 24	G>A	0,29
	rs15869	13-31871012	83395	3'UTR	A>C	0,29
FANCD2	rs781871	3-10041805	-1308	5' corriente arriba	T>A	0,35
	rs33917318	3-10042797	-316	5' corriente arriba	T>C	0,22
	rs17032268	3-10042972	-141	5' corriente arriba	C>T	0,17
	rs722509	3-10077377	34264	Intrón 19	T>A	0,20

VIII. ANEXOS

VIII.1. Anexo I: Información adicional

	rs4019784 rs3864017	3-10081408 3-10081532	38295 38419	Intrón 22 Exón 23 (Pro714Leu)	C/T G>A	- 0,23
(continuación)		Cromosoma - - Posición	Posición génica	Localización génica	Cambio nucleotídico	MAF caucásico HapMap
Gen	SNP					
FANCD2	rs9811771	3-10108949	65836	Intrón 38	A>G	0,11
	rs2272125	3-10113069	69956	Exón 42 (Leu1366Leu)	A>C	0,14
	rs7647987	3-10115696	72583	Intrón 42	A>T	0,21
	rs34305518	3-10116056	72943	3'UTR	T>A	0,22
	rs3826	3-10116065	72952	3'UTR	A>C	0,23
	rs11716842	3-10116184	73071	3'UTR	C>T	0,31
FANCE	rs2395626	6-35530732	2616	Intrón 1	T>G	0,44
	rs760782	6-35556167	28051	3' corriente abajo	G>A	0,31
FANCF	rs3740615	11-22603942	-10	5'UTR	C>T	0,13
	rs4447177	11-22602601	1331	3'UTR	A>G	0,15
FANCG	rs568300	9-35083020	-13007	5' corriente arriba	C>T	0,42
	rs2258240	9-35050955	19058	3' corriente abajo	C>T	0,20
	rs1053318	9-35046961	23052	3' corriente abajo	G>T	0,21
FANCI	rs7168941	15-87599981	11783	Intrón 2	G>A	0,07
	rs16942918	15-87605115	16917	Intrón 4	G>A	0,36
	rs8032440	15-87652595	64397	Intrón 34	G>C	0,08
	rs3087374	15-87660998	72800	3'UTR	G>T	0,08
FANCI/ BRIP1	rs8080599	21-57314371	-18834	5' corriente arriba	C>T	0,12
	rs2048718	21-57295601	-64	5' corriente arriba	C>T	0,35
	rs4988340	21-57295415	122	Intrón 1	G>A	0,21
	rs4988351	21-57212591	82946	Intrón 12	C>G	0,27
	rs9897928	21-57198812	96725	Intrón 14	T>A	0,22
	rs1557720	21-57141156	154381	Intrón 17	T>A	0,46
	rs4986765	21-57118247	177290	Exón 19 (Glu879Glu)	G>A	0,40
	rs4988357	21-57117896	177641	Intrón 19	T>A	0,33
FANCL	rs7220740	21-57095575	199962	3' corriente abajo	C>A	0,28
	rs3771224	2-58313533	8485	Intrón 1	C>T	0,24
	rs1011314	2-58276496	45522	Intrón 7	G>A	0,30
	rs848289	2-58245074	76944	Intrón 8	T>A	0,37
FANCM	rs8020533	14-44720563	45663	Intrón 15	G>A	0,07
	rs226984	14-44721431	46531	Intrón 16	A>G	0,29
	rs226981	14-44732075	57175	Intrón 20	G>A	0,55
	rs3736772	14-44735218	60318	Exón 21 (Ala1812Pro)	G>C	0,08
FANCN/ PALB1	rs8058061	16-23560521	-342	5' corriente arriba	A>G	0,06
	rs249941	16-23552239	7940	Intrón 4	C>T	0,17
	rs16940342	16-23552087	8092	Intrón 4	A>G	0,10
	rs447529	16-23540924	19255	Intrón 9	G>C	0,19
H2AX	rs8551	11-118472734	-1365	5' corriente arriba	C>T	0,46
	rs7759	11-118472501	-1132	5' corriente arriba	A>G	0,43
	rs7350	11-118470258	1111	Intrón 1	C>T	0,41
OGG1	rs159153	3-9764875	-830	5' corriente arriba	T>C	0,24
	rs125701	3-9765478	-227	5' corriente arriba	G>A	0,13
	rs3219008	3-9770543	4838	Intrón 3	A>G	0,19
	rs1052133	3-9773773	8068	Intrón 6	C>G	0,22
USP1	rs9436223	62676135	1164	Intrón 1	C>T	0,46
	rs614665	62682562	7591	Intrón 5	T/C	-
	rs583609	62689384	14413	3'UTR	T>C	0,33

Anexo I.12. Descripción de los genes de control del ciclo celular seleccionados para el estudio y las referencias bibliográficas que determinaron su elección.

Gen	Descripción	Referencias bibliográficas
BRG1	Mitotic growth and transcription activator	(Bochar <i>et al.</i> , 2000; Wang <i>et al.</i> , 2004a; Wong <i>et al.</i> , 2000)
BUB1B	Mitotic checkpoint serine/threonine-protein kinase BUB1 beta	(Lo <i>et al.</i> , 2007; Scintu <i>et al.</i> , 2007; Vaclavicek <i>et al.</i> , 2007)
c-MYC	Myc proto-oncogene protein	(Chrzan <i>et al.</i> , 2001; Ghoussaini <i>et al.</i> , 2008)
E2F1	Transcription factor E2F1	(Powers <i>et al.</i> , 2004; Rogoff and Kowalik, 2004; Vuaroqueaux <i>et al.</i> , 2007; Zacharatos <i>et al.</i> , 2004)
E2F2	Transcription factor E2F2	(Scheijen <i>et al.</i> , 2004; Timmers <i>et al.</i> , 2007)
E2F3	Transcription factor E2F3	(Timmers <i>et al.</i> , 2007)
E2F4	Transcription factor E2F4	(DuPree <i>et al.</i> , 2004)
E2F5	Transcription factor E2F5	(van den Heuvel and Dyson, 2008)
EP300	E1A-associated protein p300	(Campbell <i>et al.</i> , 2004)
MDM2	Ubiquitin-protein ligase E3 Mdm2, p53-binding protein Mdm2	(Bond <i>et al.</i> , 2004; Deb, 2003)
RB1	Retinoblastoma-associated protein	(Classon and Harlow, 2002; Lesueur <i>et al.</i> , 2006; Simin <i>et al.</i> , 2004)
RBL2/p130	Retinoblastoma-like protein 2	(Macaluso <i>et al.</i> , 2003)
STK6	Serine/threonine kinase 6, Aurora-A	(Bodvarsdottir <i>et al.</i> , 2007; Egan <i>et al.</i> , 2004; Ouchi <i>et al.</i> , 2004; Sun <i>et al.</i> , 2004)
VDR	Vitamin D receptor	(Abbas <i>et al.</i> , 2008; Guy <i>et al.</i> , 2004; Knight <i>et al.</i> , 2007; Trabert <i>et al.</i> , 2007)

Anexo I.13. Descripción de los SNPs seleccionados en los genes de control del ciclo celular.

Gen	SNP	Cromosoma - - Posición cr.	Posición génica	Localización génica	Cambio nucleotídico	MAF caucásico HapMap
BRG1	rs7935	19-10966608	34002	Exón 9 (H508H)	C>T	0,29
	rs1801514	19-10997124	64518	Exón 22 (M1035I)	C>A	0,35
BUB1B	rs3759839	15-38238880	-1650	5' corriente arriba	G>A	0,47
	rs3214012	15-38244421	3891	Intrón 1	G>T	0,28
	rs1801376	15-38265123	24593	Exón 8 (R349Q)	A>G	0,27
	rs2305653	15-38281790	41260	Intrón 12	T>C	0,35
	rs4924437	15-38307000	66470	3' corriente abajo	A>T	0,29
c-MYC	rs2070583	8-128822436	4938	3'UTR	A>G	0,01
E2F1	rs3213150	20-31735862	2009	Intrón 1	C>T	0,26
	rs2071056	20-31729174	8697	Intrón 4	T>C	0,27
	rs3213180	20-31727285	10586	3'UTR	C>G	0,13
	rs3213183	20-31726623	11248	3' corriente abajo	C>T	0,27
E2F2	rs2742976	1-23730589	-289	5' corriente arriba	G>T	0,40
	rs3218148	1-23724374	5926	Intrón 1	C>T	0,50
	rs2075995	1-23720051	10249	Exón 4 (Q226H)	C>A	0,38
	rs3218175	1-23718329	11971	Intrón 4	G>A	0,13
	rs3218211	1-23708381	21919	3'UTR	C>T	0,48
E2F3	rs4712494	6-20516898	6521	Intrón 1	T>C	0,25
	rs911361	6-20523032	12655	Intrón 1	C>T	0,48
	rs2328489	6-20526382	16005	Intrón 1	C>G	0,25
	rs4134957	6-20593843	83466	Intrón 4	T>C	0,23
	rs3763276	6-20594544	84167	Intrón 4	G>A	0,32
	rs1322842	6-20596876	86499	Intrón 6	C>T	0,42
	rs1042317	6-20600974	90597	3'UTR	C>T	0,24
E2F4	rs3729639	16-65783002	-567	5' corriente arriba	C>T	0,07
E2F5	rs4150884	8-86289031	12123	Intrón 1	A>G	0,50
	rs3808538	8-86308563	31655	Intrón 5	A>C	0,25
EP300	rs20552	22-39880986	62433	Exón 17(T1061T)	T>A	0,33
MDM2	rs1144943	12-67486012	-2235	5'upstream	G>T	0,50
	rs937282	12-67488064	-183	5' corriente arriba	G>C	0,46
	rs937283	12-67488431	184	5'UTR	G>A	0,48
	rs1470383	12-67493429	5192	Intron 2	T>C	0,16
	rs1201644	12-67500958	12711	Intron 5	C>T	0,45
	rs3730590	12-67506903	18656	Intron 7	C>T	0,25
	rs1695144	12-67510209	21962	Intron 8	A>G	0,13
	rs2589285	12-67511942	23695	Intron 8	G>T	0,37
	rs3730646	12-67518247	30000	Intrón 10	A>G	0,50
	rs3730654	12-67519173	30926	Intrón 10	T>C	0,00
RB1	rs9568029	13-47766536	-9348	5' corriente arriba	C>T	0,28
	rs1573601	13-47774358	-1526	5' corriente arriba	C>A	0,25
	rs2804094	13-47790993	15109	Intrón 2	G>T	0,03
	rs4151441	13-47799876	23992	Intrón 2	T>C	0,13
	rs520342	13-47814896	39012	Intrón 3	C>T	0,25
	rs4151474	13-47819541	43657	Intrón 4	T>A	0,02
	rs399413	13-47846570	70686	Intrón 12	G>A	0,26
	rs9568036	13-47869937	94053	Intrón 17	A>G	0,38
	rs1894255	13-47887813	111929	Intrón 17	A>G	0,24
	rs2854344	13-47895694	119810	Intrón 17	G>A	0,06
	rs1951775	13-47914064	138180	Intrón 17	G>T	0,28
	rs198575	13-47926351	150467	Intrón 18	T>C	0,03
	rs198577	13-47929538	153654	Intrón 19	C>G	0,03
	rs198580	13-47931748	155864	Intrón 19	A>G	0,01
RB1	rs198590	13-47936678	160794	Intrón 21	G>A	0,26
	rs367438	13-47947233	171349	Intrón 24	T>C	0,01
RBL2/p130	rs10748	16-52062222	36360	Exón 17 (L864L)	T>C	0,47

STK6	rs1468056	20-54399395	1363	Intrón 1	G>C	0,60
	rs2273535	20-54394948	5810	Exón 4 (I31F)	A>T	0,18
(continuación)	Cromosoma -	Posición	Localización	Cambio	MAF caucásico	
Gen	SNP	- Posición cr.	génica	nucleotídico	HapMap	
STK6	rs1047972	20-54394870	5888	Exón 4 (V57I)	G>A	0,19
	rs6014711	20-54394480	6278	Intrón 4	G>A	0,80
	rs911160	20-54390970	9788	Intrón 6	G>C	0,79
	rs1044377	20-54375312	25446	3' corriente abajo	G>T	0,23
	rs4516035	12-46586093	-1012	5' corriente arriba	T>C	0,45
VDR	rs10735810	12-46559162	25919	Exón 1 (M1T)	G>A?	0,44
	rs731236	12-46525024	60057	Exón 11 (I352I)	T>C	0,44
	rs739837	12-46524488	60593	3'UTR	T>G	0,43

Carcinogenesis vol.27 no.9 pp.1930–1937, 2006
doi:10.1093/carcin/bgl062
Advance Access publication May 5, 2006

FANCD2 associated with sporadic breast cancer risk

E.Barroso¹, R.L.Milne², L.P.Fernández¹, P.Zamora³,
J.L.Arias⁴, J.Benítez^{1,2} and G.Ribas^{1,*}

¹Human Genetics Group, ²National Genotyping Centre (CeGen), Human Cancer Genetics Programme, Spanish National Cancer Centre (CNIO), Madrid, Spain, ³Department of Oncology, La Paz Hospital, Madrid, Spain and ⁴Unit of Surgery, Monte Naranco Hospital, Oviedo, Spain

*To whom correspondence should be addressed at: Human Genetics Group, Human Cancer Genetics Program, C/Melchor Fdz Almagro, 3, E-28029, Madrid, Spain. Tel: +34 912246950; Email: gribas@cnio.es

Several components of the Fanconi anaemia (FA) family of proteins allow the formation of the DNA repair complex *foci* formed by proteins such as BRCA1/2 and RAD51. Because the genes that participate in the DNA repair pathway have been described as low-penetrance breast cancer susceptibility genes, we postulated that variants in FA genes could also be associated with sporadic breast cancer risk. We studied seven SNPs in *FANCA*, *FANCL* and *FANCD2* in a total of 897 consecutive and non-related sporadic breast cancer cases and 1033 unaffected controls from the Spanish population. We observed a statistically significant association with sporadic breast cancer for the variant rs2272125 (L1366L) located on *FANCD2* (OR per allele = 1.35; 95% C.I. 1.09–1.67; $P = 0.005$). Both haplotype and diplotype analyses confirmed this association, where one haplotype and pooled diplotypes carrying it were associated with more than 4-fold risk ($P = 0.007$ and $P = 0.006$, respectively). Screening for potential causal variants in *FANCD2* was performed, detecting one in the putative promoter region, which is located in a phylogenetically conserved motif with consensus binding sites for some transcriptional factors, suggesting a functional implication. Our data indicate that a relationship between *FANCD2* and sporadic breast cancer risk may exist.

Introduction

Breast cancer is one of the most common diseases among women in developed countries. Several breast cancer susceptibility genes have been identified, the most important being the high risk tumour suppressor genes *BRCA1* (1) and *BRCA2* (2), the latter also identified as *FANCD1* (3,4). However, under the polygenic model of complex diseases, there are also low risk genes with variants that are common in the general population and that produce a mild susceptibility to the disease. Whether these variants interact together or with other environmental and lifestyle factors in an additive or multiplicative way, is not yet clear (5–8).

Abbreviations: FA, Fanconi anaemia; DSB, double-strand break; SNP, single nucleotide polymorphism. HR, homologous recombination; LD, linkage disequilibrium; MAF, minor allele frequency.

The BRCA1 and BRCA2 (FANCD1) proteins are implicated in several cellular mechanisms, the most relevant being double-strand breaks (DSBs) by homologous recombination (HR) (9,10). Both BRCA1 and BRCA2 interact with the RAD51 protein, a key component of HR DNA damage repair, to form the nuclear repair *foci* after DNA damage (11–14). Furthermore, the FA protein family has been identified as a component of this DSB HR repair machinery (15–17). To date, 12 complementation groups have been described, for which 11 responsible FA genes have been identified: FANCA, -B, -C, -D1, -D2, -E, -F, -G, -J, -L and -M. Eight of these (FANCA, -B, -C, -E, -F, -G, -L and -M) form a multi-subunit nuclear FA complex (18–20). Some of the FA proteins are structural proteins in the FA complex. For example, FANCA acts as dimer, FANCG plays an important role in the maintenance of the complex structure (21,20) and FANCF functions as a flexible adaptor protein which is required for the assembly of this complex (22). Other FA proteins have specific functions such as FANCL, a ubiquitin ligase (23,24), whose stability and function depends on FANCB (25,26). Another example is FANCD2, a monoubiquitinated protein, which is activated by FANCL and passes from isoform FANCD2-S to FANCD2-L (27,16). FANCD2 is located downstream of the FA complex and linked with it through FANCE (20). FANCI, described previously as helicase BACH1/BRIP1 (28,29), and FANCM, a nuclease that may act as an engine that translocates the FA core complex along DNA, have been most recently identified (30).

The pathway activity that implicates FA proteins is initiated when DNA damage occurs, and ATM and ATR protein kinases function as DNA damage sensors (31,32). They phosphorylate FANCD2 which is required for the activation of an S-phase checkpoint (33,34) and for its appropriate monoubiquitination through the FA complex (35). They also phosphorylate several proteins in the FA complex (36) which are necessary for the activation of FANCL ubiquitin ligase activity. After monoubiquitination of FANCD2-S, FANCD2-L translocates to the nuclear repair complex *foci*, helped by BRCA1 in its formation and accumulation (37), where BRCA2 and RAD51 are also located (27). Thus, the monoubiquitination of FANCD2 is important because it allows its chromatin-associated *foci* assembly (38,39), either in response to DNA damage (40) during the S-phase of the normal cell cycle (41,42), or in replicative stress (43). It has therefore been proposed that FANCD2 could mediate the BRCA2–RAD51 interaction and also influence the speed and efficiency of DNA repair (17,40,44) although FANCG may also be implicated in this process (45). However, an alternative model proposes that FANCD2 functions in a DNA repair sub-pathway that is independent of DNA cross-over and sister chromatid exchange, where RAD51 focus formation depends on BRCA2, but not FANCD2 (46). It has been demonstrated that FA proteins are not essential for HR repair machinery. Furthermore, FA cells have a mild defect in the single-strand annealing pathway, which is also

FANCD2 and breast cancer risk

involved in DSB repair. Thus, FA proteins could work in other DNA damage repair pathways of HR (47).

We focused our study on the hypothesis that genes involved in the mechanism of DNA damage repair, acting similarly to tumour suppressors, would act as low susceptibility genes in the development of sporadic breast cancer, as proposed previously (48). In this study, we analysed seven genetic variants in three FA genes: two variants in *FANCA*, three in *FANCL*, and two in *FANCD2*. The first two genes encode proteins in the FA complex and the latter encodes a protein downstream of the FA complex.

Materials and methods

Subjects

This case-control study included a total of 897 consecutive and non-related sporadic breast cancer cases and 1033 healthy controls. The case samples had a mean age of diagnosis of 50 years (range 23–90 years) (Table I). They were collected from seven Spanish public hospitals: 258 (29%) from *Hospital Monte Naranjo*, in Oviedo; 164 (18%) from *Fundación Jiménez Díaz*, 132 (15%) from *Hospital La Paz* and 100 (11%) from *Hospital Clínico San Carlos*, all in Madrid; 115 (13%) from *Hospital San Pau*, in Barcelona; 100 (11%) from *Hospital de Salamanca*, in Salamanca; 28 (3%) from *Hospital de Valencia*, in Valencia. Controls (between 22 and 85 years of age) (Table I) were unaffected Spanish women collected from six centres: 462 (45%) from the Menopause Research Centre at the *Instituto Palacios*; 287 (28%) from the National Blood Transfusion Centre; 221 (21%) from *Fundación Jiménez Díaz*; 63 (6%) from *Hospital Clínico San Carlos*, all in Madrid.

Extraction of genomic DNA

Genomic DNA from cases and controls was isolated from peripheral blood lymphocytes using the MagNA Pure LC Instrument (Roche Molecular Biochemicals, Germany) following manufacturer's instructions (49), and a stock solution of DNA at 50 ng/μl was obtained, measured using the PicoGreen dsDNA Quantitation Kit (Molecular Probes, Eugene, OR).

SNP selection

Three public databases available via internet were used to collect information about single nucleotide polymorphisms (SNPs) in the three genes (*FANCA*, *FANCL* and *FANCD2*): NCBI (<http://www.ncbi.nlm.nih.gov>), Ensembl (<http://www.ensembl.org>) and HapMap (<http://www.hapmap.org>). SNPs were considered if they had a minor allele frequency (MAF) greater than or equal to 10%, the exception being putative coding SNPs with MAF \geq 5%. Information on linkage disequilibrium (LD) (blocks within chromosomes that are more probably linked and segregate together than expected at random), was used to select representative SNPs in high LD blocks. Haploview v3.11 (50) was used to select tag-SNPs (a subset of SNPs that represent variation in a gene or chromosomal region) where possible. We selected and successfully genotyped two SNPs on the 3' downstream region of *FANCA* (rs1195066 and rs2239359, spaced 13 Kb apart), three located on introns 1, 7 and 8 of *FANCL* (rs848289, rs1011314 and rs3771224, respectively, spaced over 68 Kb) and two in *FANCD2* (rs722509 and rs2272125, spaced 32 Kb apart), one on intron 19 and the other a synonymous coding change (L1366L) on exon 42. Gene structures and SNP positions are shown in Figure 1A.

Genotyping assays

Genotyping was carried out using the TaqMan platform. SNP assays were designed using Applied Biosystems Assay-by-design and Assay-on-Demand probes (Applied Biosystems, Foster City, CA, USA) (primer and probe information is provided in Supplementary Table I). PCR amplifications were carried out following the manufacturer's standardized protocol. The genotype of each sample was automatically assigned by measuring allele-specific fluorescence in the ABI Prism 7900HT Detection System, using the SDSv2.1 (Sequence Detection Software) for allele discrimination, by Applied Biosystems. Three non-DNA and nine duplicate samples were included in each plate of 384 samples used, as an internal control.

Statistical analysis

Deviations from Hardy-Weinberg equilibrium were tested using the *genhw* command in STATAv8.0 (51), applying Pearson's Chi-squared test or Fisher's Exact Test where appropriate. Differences in the MAFs between cases and controls were tested using the Chi-squared test. In order to assess associations

Table I. Age categories of control and case samples

Age (years)	Controls		Cases		Total	
	N	(%)	N	(%)	N	(%)
<40	261	(25)	209	(23)	470	(24)
40–44	61	(6)	30	(3)	91	(5)
45–49	90	(9)	45	(5)	135	(7)
50–54	156	(15)	72	(8)	228	(12)
55–59	147	(14)	74	(8)	221	(12)
60–64	91	(9)	72	(8)	163	(8)
65+	22	(2)	112	(13)	134	(7)
Unknown	205	(20)	283	(32)	488	(25)
Total	1033	(100)	897	(100)	1930	(100)

between genotypes, haplotypes (combination of variants along a chromosome) and diplotypes (haplotype combinations in individuals), and breast cancer risk, several analysis were performed. PHASEv2.0 software (52,53) was used to impute haplotypes and compare their frequency distributions in cases and controls. Unconditional logistic regression was applied using STATAv8.0 for the analysis of genotypes, haplotypes and diplotypes, the most likely imputed haplotype being assumed for the latter two. Odds ratios (OR), their 95% confidence intervals (CI) and Wald statistic *P*-values were estimated, using the most frequent genotype, haplotype and diplotype, respectively, among controls as reference. In the analysis of genotypes, dominant, recessive and multiplicative (single parameter) codominant models were assessed against the two parameter codominant model via the likelihood ratio test, and where the latter was not the best fit (nominal *P*-value greater than 0.05), the best fitting model was determined by parsimony. Age was included as a categorical variable (<40, 40–44, 45–49, 50–54, 55–59, 60–64, \geq 65) (Table I). Analysis of haplotypes was repeated using the *haplo.stats* library implemented in R which compares haplotype frequencies in cases and controls in an unbiased way by including haplotype uncertainty in the estimation of ORs. For all statistical tests, results were considered significant for two-sided *P*-values less than 0.05.

Screening for functional variants in the FANCD2 gene

Haploview v3.11 was used to graphically analyse the extent of LD in the *FANCD2* region (Figure 1B). Selected sequences in *FANCD2* (promoter region and nine exons) were further analysed by PCR, dHPLC and sequencing in samples carrying haplotype H2 in the *FANCD2* gene. Primers in flanking exons, intron-exon boundaries and the promoter region were designed using Primer3 software (<http://frodo.wi.mit.edu/cgi-bin/primer3/primer3-www.cgi>) (54), based on genomic information collected from NCBI (<http://www.ncbi.nlm.nih.gov>). PCR products, obtained under standardized conditions, were analysed by dHPLC using the WAVE DNA-fragment analysis system (Transgenomics Limited, Crewe, UK), and were determined by the WAVEMAKER software from Transgenomics (PCR primers and conditions, and dHPLC elution conditions are shown in Supplementary Table II). Those DNAs with heterozygous peaks detected by dHPLC analysis were purified using the High Pure PCR Product Purification Kit (Roche) and sequenced with the ABI Prism Big DyeTM System and ABI 3700 sequencer (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's protocol. Additional validation for the identified variants ss49839682, rs17032268, rs9811771 and rs7647987 in *FANCD2* was carried out by dHPLC analysis (conditions shown in Supplementary Table II) in 95 case and 95 control non-H2 carriers. A sample of 551 case and 559 control non-H2 carriers were further genotyped using TaqMan probes (Supplementary Table I) in order to validate the results obtained for ss49839682 and rs9811771. *FANCD2* variants were assessed for functional importance using PupaSNP (<http://pupasnp.bioinfo.ocha.fib.es>), which synthesizes information from: Ensembl v25.34e, dbSNP 122, HapMap release #12, Match v2.2, Transfac Professional v8.3, Haploview v3.0 and PMUT (55). Transfac Professional v8.3 was also applied to screen for transcription factor consensus binding sites using rVista v2.0 (<http://rvista.dcode.org>) (56). Phylogenetic conservation analysis was performed using ECR Browser (<http://ecrbrowser.dcode.org>) (57). Phylogenetic sequence comparisons and construction of phylogenetic trees among mammalian organisms were carried out using Tcoffee (<http://igs-server.cnrs-mrs.fr/Tcoffee/tcoffee.cgi/index.cgi>) (58) (see Figures 2A and B and Supplementary Figure 1).

E.Barroso *et al.*

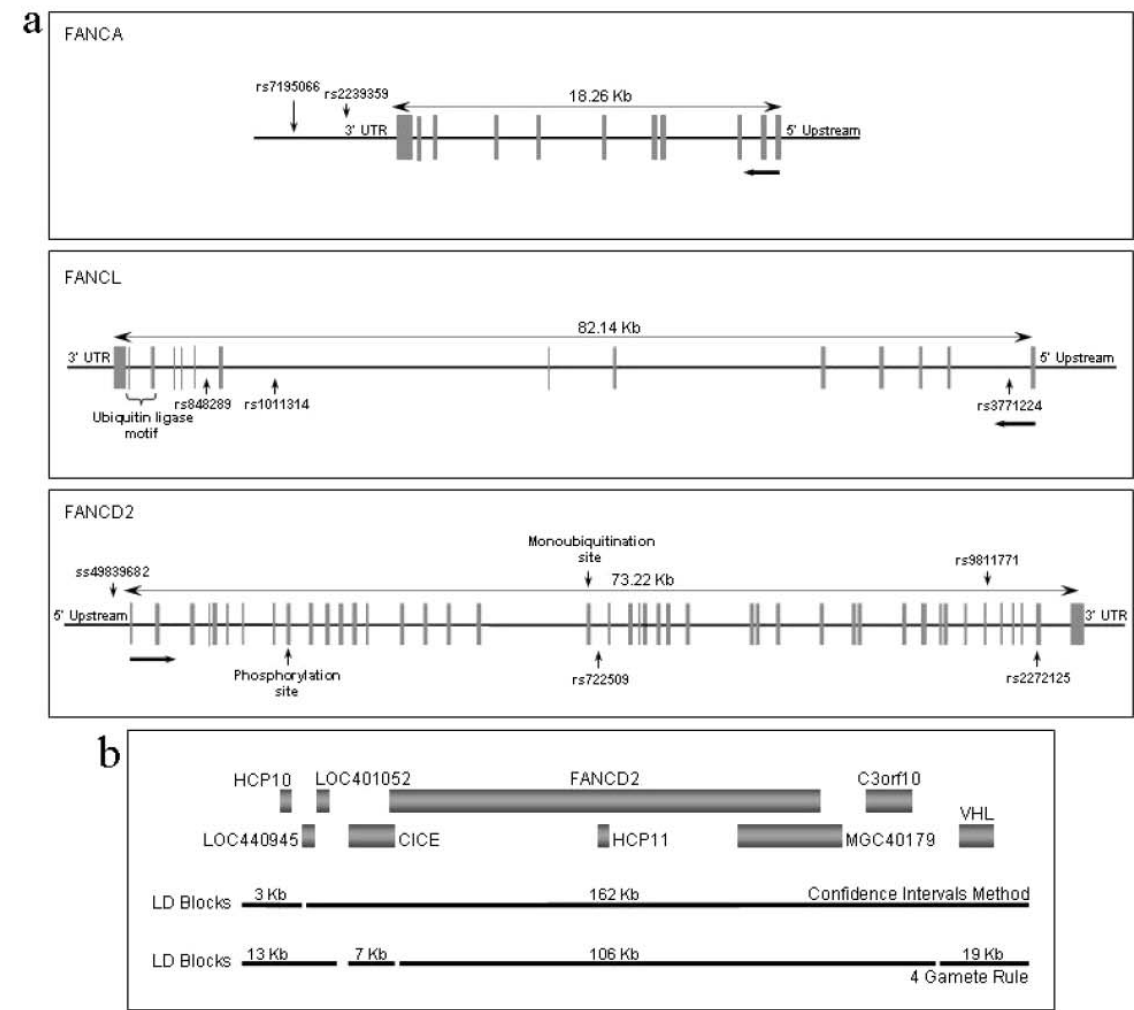


Fig. 1. (A) Gene structure and SNP location. Schematic diagram of the *FANCA*, *FANCL* and *FANCD2* genes. Grey boxes represent exons and solid lines represent untranslated and intron regions, with 5' upstream and 3'-UTR regions noted. Fine and arrowed lines represent gene lengths and solid arrows represent 5'-3' gene orientation. Functional motifs and assayed SNP locations are noted. (B) LD blocks located in the *FANCD2* gene region. Schematic diagram of *FANCD2* and surrounding genes located in the same LD block. Grey boxes represent genes. Solid lines represent LD blocks obtained by the Confidence Intervals Method or the 4 Gamete Rule.

Results

Genotyping

Genotyping call rates were over 97%, and concordance rates for non-DNA and duplicate samples were 100%, for all SNPs genotyped (data are shown in Table II). All seven SNPs complied with Hardy-Weinberg equilibrium among controls.

Analysis of genotypes

A higher MAF was observed among cases than controls for rs2272125 in *FANCD2* (0.13 versus 0.10, respectively, $P = 0.005$). The other six polymorphisms tested did not show any evidence of differences in the allelic frequencies (data not shown). Comparison of genotype frequencies suggested that the risk of developing sporadic breast cancer

associated with rs2272125 followed a codominant model, with the estimated OR for rare CC homozygotes relative to common AA homozygotes (OR = 2.77; 95% CI = 1.08–7.13; $P = 0.03$) more than double that for heterozygotes (OR = 1.28; 95% CI = 1.02–1.62; $P = 0.04$). The best fitting model was a single parameter codominant one, with an associated multiplicative per-allele OR of 1.35 (95% CI = 1.09–1.67, $P = 0.005$). This association was consistently observed after adjusting for age (Table I) (OR = 1.32, $P = 0.03$). We did not find any other associations among the remaining SNPs in this study. Results for all genotyped SNPs are summarized in Table II.

The program Structure v.2.0 was applied to a set of 28 unlinked biallelic markers genotyped in a random subset of 163 cases and 180 controls, finding no evidence of population substructure, with an estimated posterior probability of

FANCD2 and breast cancer risk

Table II. Statistical analysis of genotypes in *FANCA*, *FANCL* and *FANCD2*

Gene	SNP name	Spanish control MAF	Genotype	Controls		Cases		Call rate (%)	OR (95% CI)	P-value
				N	(%)	N	(%)			
<i>FANCA</i>	rs7195066	0.29	CC	269	(49)	274	(50)	99.5	1.00 (Reference)	0.65
			CT	239	(43)	230	(42)		0.94 (0.74–1.21)	
			TT	44	(8)	45	(8)		1.00 (0.64–1.57)	
	rs2239359	0.31	CC	268	(49)	251	(46)	99.3	1.00 (Reference)	0.47
			CT	223	(40)	229	(42)		1.10 (0.85–1.41)	
			TT	60	(11)	64	(12)		1.14 (0.77–1.69)	
<i>FANCL</i>	rs848289	0.30	TT	267	(49)	252	(47)	98.9	1.00 (Reference)	0.51
			TA	236	(43)	238	(44)		1.07 (0.83–1.37)	
			AA	47	(9)	51	(9)		1.15 (0.75–1.77)	
	rs1011314	0.30	GG	276	(50)	257	(47)	99.2	1.00 (Reference)	0.51
			GA	223	(40)	226	(42)		1.09 (0.85–1.40)	
			AA	52	(9)	60	(11)		1.24 (0.42–1.86)	
<i>FANCD2</i>	rs3771224	0.30	CC	262	(48)	240	(45)	97.1	1.00 (Reference)	0.40
			CT	233	(43)	238	(45)		1.12 (0.87–1.43)	
			TT	46	(9)	52	(10)		1.23 (0.80–1.90)	
	rs722509	0.11	TT	739	(79)	650	(77)	97.1	1.00 (Reference)	0.20
			TA	181	(19)	185	(22)		1.16 (0.92–1.46)	
			AA	11	(1)	12	(1)		1.24 (0.54–2.83)	
	rs2272125	0.10	AA	675	(80)	649	(75)	97.4	1.00 (Reference)	0.04
			AC	162	(19)	200	(23)		1.28 (1.02–1.62)	
			CC	6	(1)	16	(2)		2.77 (1.08–7.13)	

Genotype counts, call rates and frequencies in cases and controls with ORs and P-values for tests of association between genotypes and sporadic breast cancer.
OR, odds ratio; CI, confidence interval.

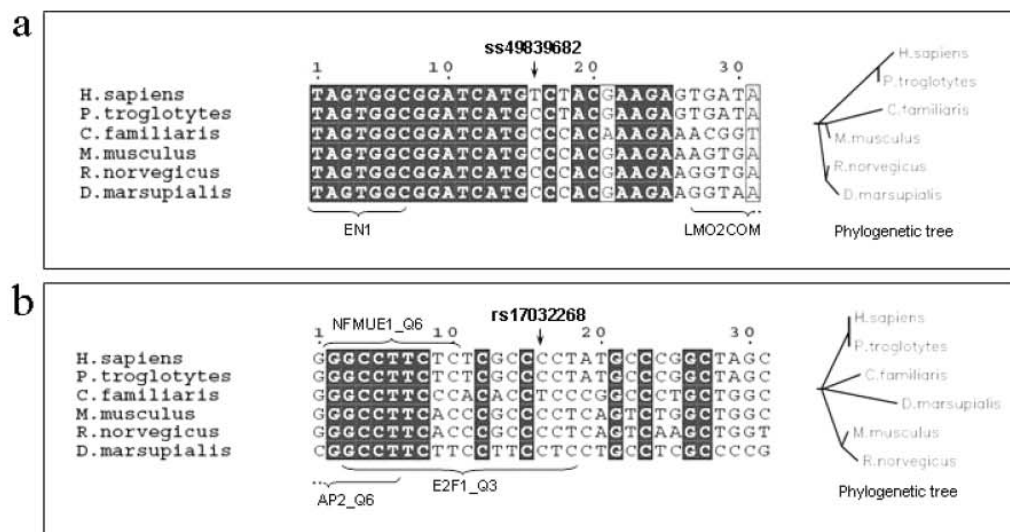


Fig. 2. Phylogenetic analyses of ss49839682 (A) and rs17032268 (B) sequence conservation in mammalian organisms. Sequence comparison of surrounding variants and transcription factor consensus binding sites. Phylogenetic trees as measures of divergence are noted. Grey boxes represent highly conserved sequences. SNP sites and transcription factor binding sites are noted.

approximately one for K (number of clusters) = 1, and minimum and maximum ancestry coefficients of 0.45 and 0.55, respectively for K = 2, consistent across independent runs (R.L. Milne unpublished data).

Analysis of haplotypes and diplotypes

We compared the imputed haplotypes and diplotypes formed by the SNPs in *FANCD2* between cases and controls and the

results are presented in Table III. The most common haplotype (H1) found in both cases and controls carried the more common alleles 'T' in rs722509 and 'A' in rs2272125. Comparison of haplotype frequencies between cases and controls suggested that one of the four haplotypes was associated with the risk of breast cancer. This haplotype (H2) carried the combination 'TC' and the estimated OR relative to H1 was 4.46 (95% CI = 1.51–13.21; P = 0.007) assuming the most

E.Barroso *et al.***Table III.** Statistical analysis of haplotypes (a) and diplotypes (b) in the *FANCD2* gene. Bold typeface indicate statistically significant haplotype and diplotype analyses.

	Controls		Cases		OR (95% CI)	P-value
	N	(%)	N	(%)		
(a) Haplotype ^a						
H1	1328	(89)	1340	(87)	1.00 (Reference)	
H2	4	(0)	18	(1)	4.46 (1.51–13.21)	0.007
H3	4	(0)	3	(0)	0.74 (0.17–3.33)	0.70
H4	150	(10)	185	(12)	1.22 (0.97–1.54)	0.09
(b) Diplotype ^b						
D1-1	592	(80)	581	(75)	1.00 (Reference)	
D1-2	4	(1)	14	(2)	3.57 (1.17–10.90)	0.03
D1-3	2	(0)	3	(0)	1.53 (0.25–9.18)	0.64
D1-4	138	(19)	161	(21)	1.19 (0.92–1.53)	0.18
D2-4	0	(0)	4	(1)	—	—
D3-3	1	(0)	0	(0)	—	—
D4-4	6	(1)	10	(1)	1.70 (0.61–4.70)	0.31
D2-	4	(1)	18	(2)	4.59 (1.54–13.63)	0.006

Counts and frequencies in cases and controls with ORs and P-values for tests of association with sporadic breast cancer.

OR, odds ratio; CI, confidence interval.

D2- represent people carrying at least one haplotype H2 (D1-2 and D2-4).

^aH1 = TA; H2 = TC; H3 = AA; H4 = AC.

^bD1-1 = H1/H1; D1-2 = H1/H2; D1-3 = H1/H3; D1-4 = H1/H4; D2-4 = H2/H4; D3-3 = H3/H3; D4-4 = H4/H4.

likely haplotypes were actually observed, and 4.81 (95% CI = 1.63–14.18; $P = 0.004$) allowing for haplotype uncertainty [Table III (a)].

Analysis of diplotypes found that carrying H1 and H2 (diplotype D1-2) was associated with increased risk of breast cancer relative to D1-1 (OR 3.57; 95% CI = 1.17–10.90; $P = 0.03$). Moreover, only four women carried H2 in a diplotype other than D1-2, all in combination with H4 (D2-4) and all four were cases. Pooling all carriers of H2 (D1-2 and D2-4) confirmed that they were at increased risk (OR = 4.59; 95% CI = 1.54–13.63; $P = 0.006$) [Table III (b)].

Screening for putative causal mutations in *FANCD2* risk haplotype carriers

rs2272125 is unlikely to directly cause the observed increased breast cancer risk because it is a synonymous coding SNP. Furthermore, we evaluated its potential implication in exon-intron splicing and found no exonic splicing enhancer (ESE) motifs in that region of *FANCD2*. The observed association might be due to a different causal variant in the same LD block as rs2272125, which extends over the whole region including *FANCD2* (over 100kb, see Figure 1B). Because it was unclear where the causative variation could be located, we searched for putative causal variants in the region among the carriers of the risk haplotype H2 (D1-2 and D2-4). We studied the described functional regions in the *FANCD2* protein by DNA sequencing: the phosphorylation motif, located in exon 9, and the monoubiquitination motif, located in exon 19 (Figure 1A). We did not detect any variation from the reference sequence used (NC_000004).

We also sought to characterize the region surrounding rs2272125, (from exon 38 to exon 43), as well as the promoter (1500 bp upstream of ATG) and 3'-UTR regions, the latter two as putative functional domains. Three variants were found in the putative promoter region: rs781871 (MAF not described),

rs17032268 (NCBI MAF = 0.22) and a novel SNP ss49839682 (–544 bp). A transitional change A > G in intron 38, 13 bp from exon 38 (IV38+13), was identified in the public databases as rs9811771, with unknown MAF. Finally, four SNPs in the 3'-UTR region were detected, one (rs7647987) located before the polyA sequence (NCBI MAF = 0.12), and three others located after the polyA sequence: rs3826, rs11716842, and a novel polymorphism ss49839681 (+421 bp from the stop codon), all with unknown MAF (Table IV).

All these variants detected in the H2 carriers were in high LD and, according to PHASEv2.0, the variant-haplotype VH2 (CTGGAG) was observed (see areas highlighted in grey in Table IV). Four individuals carried sections of VH2 and a fifth carried the entire haplotype, together representing 30% of the H2 carriers.

We genotyped four of these eight variants in 190 non-carriers of H2 (95 cases and 95 controls) in order to investigate whether any appeared accumulated in H2 carriers. These variants were: ss49839682, rs17032268, rs9811771 and rs7647987, chosen because our hypothesis was focused on detecting variants with functional implications that also shared the main part of VH2 (Table IV). Only two out of the four variants selected (ss49839682 and rs9811771) appeared to be more frequent in H2 carriers than non-carriers although neither was statistically significant (OR = 2.04; 95% CI = 0.73–5.66; $P = 0.2$; and OR = 1.68; 95% CI = 0.61–4.62; $P = 0.3$, respectively). We subsequently genotyped a sample of 1100 DNAs for these two SNPs, but did not observe statistical significance, although the tendency of the minor allele accumulating in H2 carriers was also observed (Table V).

Analysis of sequence conservation was performed for these four additional *FANCD2* variants in several mammal organisms. Only ss49839682 and rs17032268 (and their surrounding regions) have been phylogenetically maintained in mammals. Furthermore, these two variants located in the promoter region contain several consensus motifs for transcriptional factors (Figure 2A and B, respectively). The variant rs9811771, located near the exon-intron splicing site, was only weakly conserved in *Pan troglodytes* and *Canis familiaris* and rs7647987, located at the 3'polyA site, was not conserved at all (see Supplementary Figures 1A and B, respectively).

Discussion

Previous studies have demonstrated that several genes involved in DNA repair, other than *BRCA1* or *BRCA2*, are related to breast cancer susceptibility (59–63). *BRCA2* has been identified as *FANCD1* (3), and several interactions between both *BRCA1* and *BRCA2* and FA proteins have been described (64,45,40,65). It is therefore plausible that alterations in other FA genes might be implicated in breast cancer risk (66). It is known that alterations in FA genes are linked to several types of cancer, included breast cancer, in the general (non-FA) population (67,68). For example, the absence of FANCF by aberrant promoter methylation (69,70), and *BACH1/FANCF* mutations (71) have both been found in breast cancer cases. Furthermore, one study found that *FANCD2* is more often expressed in high proliferation rate tissues, such as mammary glands (72), and mouse model studies have shown that a germline disruption of the murine *Fancd2* gene results in ovarian and epithelial breast cancers (73,74). However, despite the evidence in the literature about

VIII. ANEXOS

VIII.2. Anexo II: Publicaciones originadas por la tesis

FANCD2 and breast cancer risk

Table IV. Variants found by dHPLC and sequencing in samples carrying haplotype H2.

Sample	Promoter			Intron 38		Exon 43		
	rs781871	ss49839682	rs17032268	rs9811771	rs7647987	ss49839681	rs3826	rs11716842
Major allele	T	T	C	A	A	T	T	C
Case-1	T/T	T/C	C/T	A/A	A/A	—	—	—
Case-2	T/A	T/T	C/C	A/A	A/A	—	—	—
Case-3	T/T	T/C	C/T	A/G	A/G	T/A	T/G	C/T
Case-4	T/A	T/T	C/C	A/A	A/A	T/A	T/T	C/T
Case-5	T/T	T/T	C/C	A/A	A/A	T/T	T/T	C/C
Case-6	—	T/T	C/C	A/A	A/A	T/T	T/T	C/C
Case-7	T/T	T/T	C/C	A/G	A/A	T/T	T/T	C/C
Case-8	T/T	T/T	C/C	A/A	A/A	—	—	—
Case-9	—	—	—	A/A	A/A	—	—	—
Case-10	—	T/T	C/C	A/A	A/A	T/T	T/T	C/C
Case-11	—	—	—	A/A	A/A	—	—	—
Case-12	T/A	T/T	C/C	A/A	A/G	T/T	T/T	C/C
Case-13	T/T	T/C	C/T	A/G	A/G	T/T	T/T	C/C
Case-14	T/T	T/T	C/C	A/G	A/G	T/A	T/G	C/C
Control-1	T/T	T/C	C/C	A/A	A/G	T/T	T/T	C/C
Control-2	T/T	T/C	C/T	A/G	A/G	—	—	—
Control-3	T/T	T/T	C/C	A/A	A/A	T/A	T/T	C/C

—, unknown genotype, not amplified.

Samples sharing sections of the same haplotype (VH2) highlighted in grey.

Variants between the dashed lines were selected for further study.

Table V. Observed allele frequencies for four putative functional SNPs, by H2 carrier status

Variant	H2 carriers		190 H2 non-carriers		1100 H2 non-carriers	
	N_{ma}/N_a	MAF	N_{ma}/N_a	MAF	N_{ma}/N_a	MAF
ss49839682	5/30	0.17*	34/380	0.09*	248/2166	0.11*
rs17032268	4/26	0.13	34/346	0.10	—	—
rs9811771	5/34	0.15**	35/376	0.09**	265/2154	0.12**
rs7647987	6/28	0.18	54/298	0.18	—	—

N_{ma} , observed number of minor alleles; N_a , total number of alleles; MAF, minor allele frequency.

* P (comparison of MAFs) = 0.2 and 0.4, using 190 and 1083 non-H2 carrier samples, respectively.

** P (comparison of MAFs) = 0.3 and 0.7, using 188 or 1077 non-H2 carrier samples, respectively.

the potential susceptibility to breast and other cancers associated with the FA family (75), there have been few studies investigating these associations (76).

In this study, we have investigated associations between genetic variants in three genes belonging to the FA family and sporadic breast cancer in the Spanish population. We genotyped seven SNPs in *FANCA*, *FANCL* and *FANCD2*. Whereas no associations with breast cancer risk were found for SNPs in *FANCA* and *FANCL*, we were able to identify an increased risk associated with the minor 'C' allele of rs2272125 in *FANCD2* (OR = 1.35 per allele) (Table II). Both haplotype and diplotype analyses confirmed this association, with carriers of at least one copy of the H2 haplotype at an estimated greater than 4-fold risk of breast cancer ($P = 0.007$ and $P = 0.006$, respectively) (Table III).

Despite being located in a coding region, the variant rs2272125 is of a synonymous nature and it is not located in any ESE motif. A likely explanation for this association is therefore that there is a causative functional change located in close proximity or in LD with this SNP. The region of high

LD that includes rs2272125 is over 100 Kb and spans the entire *FANCD2* gene. Applying two methods of LD inference, the Confidence Intervals Method (77) and the 4 Gametes Rule (78), it was found that this LD block contains *FANCD2* and three other genes of unknown function. *VHL*, a gene close to *FANCD2* which has been implicated in some cancers (79,80), was excluded because it was not included in the LD block according to the latter method (Figure 1B). We therefore focused on *FANCD2*, our hypothesis being that this gene could carry the putative causal variant.

In-depth analysis of several regions of the *FANCD2* gene among H2 carriers allowed us to detect another eight variants, some of them described previously in the SNP databases, but others identified for the first time (ss49839682 and ss49839681) (Table IV). Four of these variants could have functional implications because of their location (promoter, exon-intron splicing site, and 3'-UTR regions). In addition, they seemed to form a variant-haplotype (VH2) within H2 carriers (Table IV). We therefore investigated how frequent these variants were in people not carrying the H2 haplotype. One of the two variants located in the likely promoter region fell in a strongly phylogenetically conserved motif with consensus binding sites for several transcriptional factors (Figure 2A), which made us speculate that it may have functional implications; however, we have been unable to validate statistically this hypothesis (Table V). It is also possible that other variants along *FANCD2* or the immediate surrounding region could be responsible for the increased risk associated with this gene.

Another, less likely possibility, is that the observed association could be due to confounding by unmeasured factors, such as race. Although population stratification was tested for and not detected, this we done using only a subset of the sample and using only 28 markers. Population stratification cannot therefore be entirely discarded, although it appears unlikely to have influenced the results of this study. Alternatively, the observed association could be due to chance. Replication of these findings using samples from other populations is required to address these two concerns.

E.Barroso *et al.*

In conclusion, we found that a synonymous variant (rs2272125) and haplotype (H2) in the *FANCD2* gene are associated with sporadic breast cancer. This is the first study to identify this association. We have also identified a putative functional variant (ss49839682) in the same high LD block, but were not able to conclude that this is causal. Further studies will be required to clarify the full implication of the *FANCD2* protein in sporadic breast cancer risk.

Supplementary material

Supplementary material is available online at <http://www.carcin.oupjournals.org/>.

Acknowledgments

This study was supported by the grant BF12003-03852. EB and LPF were funded by the *Comunidad Autónoma de Madrid* and by the Spanish Ministry of Science and Technology (MCT), respectively. We would like to thank Álvaro Ruibal (University Hospital, Santiago de Compostela), and Santiago Palacios (*Instituto Palacios*, Madrid) for the use of samples of cases and controls. We would also like to thank Raquel Rodríguez-López for her contribution to the beginning of this work, and Fátima Mercadillo, Alicia Barroso, Emilio González, Jesús López, Victoria Fernández and Rocío Letón for their expert technical skills.

Conflict of Interest Statement: None declared.

References

- Miki,Y., Swensen,J., Shattuck-Eidens,D. *et al.* (1994) A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science*, **266**, 66–71.
- Wooster,R., Bignell,G., Lancaster,J., Swift,S., Seal,S., Mangion,J., Collins,N., Gregory,S., Gumbs,C. and Micklem,G. (1995) Identification of the breast cancer susceptibility gene BRCA2. *Nature*, **378**, 789–792.
- Howlett,N.G., Taniguchi,T., Olson,S. *et al.* (2002) Biallelic inactivation of BRCA2 in Fanconi anemia. *Science*, **297**, 606–609.
- Stewart,G. and Elledge,S.J. (2002) The two faces of BRCA2, a FANCD2 discovery. *Mol. Cell*, **10**, 2–4.
- Risch,N. (2001) The genetic epidemiology of cancer: interpreting family and twin studies and their implications for molecular genetic approaches. *Cancer Epidemiol. Biomarkers Prev.*, **10**, 733–741.
- Antoniou,A.C., Pharoah,P.D., McMullan,G., Day,N.E., Ponder,B.A. and Easton,D. (2001) Evidence for further breast cancer susceptibility genes in addition to BRCA1 and BRCA2 in a population-based study. *Genet. Epidemiol.*, **21**, 1–18.
- Antoniou,A.C. and Easton,D.F. (2003) Polygenic inheritance of breast cancer: Implications for design of association studies. *Genet. Epidemiol.*, **25**, 190–202.
- Comings,D.E., Gade-Andavolu,R., Cone,L.A., Muhleman,D. and MacMurray,J.P. (2003) A multigene test for the risk of sporadic breast carcinoma. *Cancer*, **97**, 2160–2170.
- Liang,F., Han,M., Romanienko,P.J. and Jasin,M. (1998) Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *Proc. Natl Acad. Sci. USA*, **95**, 5172–5177.
- Moynahan,M.E., Chiu,J.W., Koller,B.H. and Jasin,M. (1999) Brca1 controls homology-directed DNA repair. *Mol. Cell*, **4**, 511–518.
- Marmorstein,L.Y., Ouchi,T. and Aaronson,S.A. (1998) The BRCA2 gene product functionally interacts with p53 and RAD51. *Proc. Natl Acad. Sci. USA*, **95**, 13869–13874.
- Yuan,S.S., Lee,S.Y., Chen,G., Song,M., Tomlinson,G.E. and Lee,E.Y. (1999) BRCA2 is required for ionizing radiation-induced assembly of Rad51 complex *in vivo*. *Cancer Res.*, **59**, 3547–3551.
- Moynahan,M.E., Pierce,A.J. and Jasin,M. (2001) BRCA2 is required for homology-directed repair of chromosomal breaks. *Mol. Cell*, **7**, 263–272.
- Davies,A.A., Masson,J.Y., McIlwraith,M.J., Stasiak,A.Z., Stasiak,A., Venkitaraman,A.R. and West,S.C. (2001) Role of BRCA2 in control of the RAD51 recombination and DNA repair protein. *Mol. Cell*, **7**, 273–282.
- Digweed,M., Rothe,S., Demuth,I., Scholz,R., Schindler,D., Stumm,M., Grompe,M., Jordan,A. and Sperling,K. (2002) Attenuation of the formation of DNA-repair foci containing RAD51 in Fanconi anaemia. *Carcinogenesis*, **23**, 1121–1126.
- D'Andrea,A.D. and Grompe,M. (2003) The Fanconi anaemia/BRCA pathway. *Nat. Rev. Cancer*, **3**, 23–34.
- Wang,X. and D'Andrea,A.D. (2004) The interplay of Fanconi anemia proteins in the DNA damage response. *DNA Repair (Amst)*, **3**, 1063–1069.
- Joenje,H., Oostra,A.B., Wijker,M. *et al.* (1997) Evidence for at least eight Fanconi anemia genes. *Am. J. Hum. Genet.*, **61**, 940–944.
- Medhurst,A.L., Huber,P.A., Waisfisz,Q., de Winter,J.P. and Mathew,C.G. (2001) Direct interactions of the five known Fanconi anemia proteins suggest a common functional pathway. *Hum. Mol. Genet.*, **10**, 423–429.
- Gordon,S.M. and Buchwald,M. (2003) Fanconi anemia protein complex: mapping protein interactions in the yeast 2- and 3-hybrid systems. *Blood*, **102**, 136–141.
- Nakanishi,K., Moran,A., Hays,T., Kuang,Y., Fox,E., Garneau,D., de Oca,R.M., Grompe,M. and D'Andrea,A.D. (2001) Functional analysis of patient-derived mutations in the Fanconi anemia gene, FANCG/XRCC9. *Exp. Hematol.*, **29**, 842–849.
- Leveille,F., Blom,E., Medhurst,A.L. *et al.* (2004) The Fanconi anemia gene product FANCF is a flexible adaptor protein. *J. Biol. Chem.*, **279**, 39421–39430.
- Meetei,A.R., de Winter,J.P., Medhurst,A.L. *et al.* (2003) A novel ubiquitin ligase is deficient in Fanconi anemia. *Nat. Genet.*, **35**, 165–170.
- Meetei,A.R., Yan,Z. and Wang,W. (2004) FANCL replaces BRCA1 as the likely ubiquitin ligase responsible for FANCD2 monoubiquitination. *Cell Cycle*, **3**, 179–181.
- Meetei,A.R., Levitus,M., Xue,Y. *et al.* (2004) X-linked inheritance of Fanconi anemia complementation group B. *Nat. Genet.*, **36**, 1219–1224.
- Fei,P., Yin,J. and Wang,W. (2005) New advances in the DNA damage response network of Fanconi anemia and BRCA proteins. FAAP95 replaces BRCA2 as the true FANCB protein. *Cell Cycle*, **4**, 80–86.
- García-Higuera,I., Taniguchi,T., Ganesan,S., Meyn,M.S., Timmers,C., Hejna,J., Grompe,M. and D'Andrea,A.D. (2001) Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway. *Mol. Cell*, **7**, 249–262.
- Litman,R., Peng,M., Jin,Z., Zhang,F., Zhang,J., Powell,S., Andreassen,P.R. and Cantor,S.B. (2005) BACH1 is critical for homologous recombination and appears to be the Fanconi anemia gene product FANCI. *Cancer Cell*, **8**, 255–265.
- Bridge,W.L., Vandenberg,C.J., Franklin,R.J. and Hiom,K. (2005) The BRIP1 helicase functions independently of BRCA1 in the Fanconi anemia pathway for DNA crosslink repair. *Nat. Genet.*, **37**, 953–957.
- Meetei,A.R., Medhurst,A.L., Ling,C. *et al.* (2005) A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M. *Nat. Genet.*, **37**, 958–963.
- Morrison,C., Sonoda,E., Takao,N., Shinohara,A., Yamamoto,K. and Takeda,S. (2000) The controlling role of ATM in homologous recombinational repair of DNA damage. *EMBO J.*, **19**, 463–471.
- Cliby,W.A., Roberts,C.J., Cimprich,K.A., Stringer,C.M., Lamb,J.R., Schreiber,S.L. and Friend,S.H. (1998) Overexpression of a kinase-inactive ATR protein causes sensitivity to DNA-damaging agents and defects in cell cycle checkpoints. *EMBO J.*, **17**, 159–169.
- Grompe,M. (2002) FANCD2: a branch-point in DNA damage response? *Nat. Med.*, **8**, 555–556.
- Taniguchi,T., García-Higuera,I., Xu,B., Andreassen,P.R., Gregory,R.C., Kim,S.T., Lane,W.S., Kastan,M.B. and D'Andrea,A.D. (2002) Convergence of the fanconi anemia and ataxia telangiectasia signaling pathways. *Cell*, **109**, 459–472.
- Andreassen,P.R., D'Andrea,A.D. and Taniguchi,T. (2004) ATR couples FANCD2 monoubiquitination to the DNA-damage response. *Genes Dev.*, **18**, 1958–1963.
- Venkitaraman,A.R. (2004) Tracing the network connecting BRCA and Fanconi anaemia proteins. *Nat. Rev. Cancer*, **4**, 266–276.
- Vandenberg,C.J., Gergely,F., Ong,C.Y., Pace,P., Mallery,D.L., Hiom,K. and Patel,K.J. (2003) BRCA1-independent ubiquitination of FANCD2. *Mol. Cell*, **12**, 247–254.
- Montes de Oca,R., Andreassen,P.R., Margossian,S.P., Gregory,R.C., Taniguchi,T., Wang,X., Houghtaling,S., Grompe,M. and D'Andrea,A.D. (2005) Regulated interaction of the Fanconi anemia protein, FANCD2, with chromatin. *Blood*, **105**, 1003–1009.
- Park,W.H., Margossian,S., Horwitz,A.A., Simons,A.M., D'Andrea,A.D. and Parvin,J.D. (2005) Direct DNA binding activity of the Fanconi anemia D2 protein. *J. Biol. Chem.*, **280**, 23593–23598.
- Hussain,S., Wilson,J.B., Medhurst,A.L. *et al.* (2004) Direct interaction of FANCD2 with BRCA2 in DNA damage response pathways. *Hum. Mol. Genet.*, **13**, 1241–1248.

41. Taniguchi, T., Garcia-Higuera, I., Andreassen, P.R., Gregory, R.C., Grompe, M. and D'Andrea, A.D. (2002) S-phase-specific interaction of the Fanconi anemia protein, FANCD2, with BRCA1 and RAD51. *Blood*, **100**, 2414–2420.
42. Rothfuss, A. and Grompe, M. (2004) Repair kinetics of genomic interstrand DNA cross-links: evidence for DNA double-strand break-dependent activation of the Fanconi anemia/BRCA pathway. *Mol. Cell. Biol.*, **24**, 123–134.
43. Howlett, N.G., Taniguchi, T., Durkin, S.G., D'Andrea, A.D. and Glover, T.W. (2005) The Fanconi anemia pathway is required for the DNA replication stress response and for the regulation of common fragile site stability. *Hum. Mol. Genet.*, **14**, 693–701.
44. Wang, X., Andreassen, P.R. and D'Andrea, A.D. (2004) Functional interaction of monoubiquitinated FANCD2 and BRCA2/FANCD1 in chromatin. *Mol. Cell. Biol.*, **24**, 5850–5862.
45. Hussain, S., Witt, E., Huber, P.A., Medhurst, A.L., Ashworth, A. and Mathew, C.G. (2003) Direct interaction of the Fanconi anaemia protein FANCD2 with BRCA2/FANCD1. *Hum. Mol. Genet.*, **12**, 2503–2510.
46. Ohashi, A., Zdzienicka, M.Z., Chen, J. and Couch, F.J. (2005) Fanconi anemia complementation group D2 (FANCD2) functions independently of BRCA2- and RAD51-associated homologous recombination in response to DNA damage. *J. Biol. Chem.*, **280**, 14877–14883.
47. Nakanishi, K., Yang, Y.G., Pierce, A.J., Taniguchi, T., Digweed, M., D'Andrea, A.D., Wang, Z.Q. and Jasin, M. (2005) Human Fanconi anemia monoubiquitination pathway promotes homologous DNA repair. *Proc. Natl Acad. Sci. USA*, **102**, 1110–1115.
48. Kuschel, B., Auranen, A., McBride, S. et al. (2002) Variants in DNA double-strand break repair genes and breast cancer susceptibility. *Hum. Mol. Genet.*, **11**, 1399–1407.
49. Kessler, H.H., Muhlbauer, G., Stelzl, E., Daghofer, E., Santner, B.I. and Marth, E. (2001) Fully automated nucleic acid extraction: MagNA Pure LC. *Clin. Chem.*, **47**, 1124–1126.
50. Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.
51. StataCorp. (2003) *Stata Statistical Software: Release 8.0TX*: Stata Corporation, College Station.
52. Stephens, M., Smith, N.J. and Donnelly, P. (2001) A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.*, **68**, 978–989.
53. Stephens, M. and Donnelly, P. (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet.*, **73**, 1162–1169.
54. Rozen, S. and Skaletsky, H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.*, **132**, 365–386.
55. Conde, L., Vaquerizas, J.M., Santoyo, J., Al-Shahrour, F., Ruiz-Llorente, S., Robledo, M. and Dopazo, J. (2004) PupaSNP Finder: a web tool for finding SNPs with putative effect at transcriptional level. *Nucleic Acids Res.*, **32**, W242–W248.
56. Loots, G.G. and Ovcharenko, I. (2004) rVISTA 2.0: evolutionary analysis of transcription factor binding sites. *Nucleic Acids Res.*, **32**, W217–W221.
57. Ovcharenko, I., Nobrega, M.A., Loots, G.G. and Stubbs, L. (2004) ECR Browser: a tool for visualizing and accessing data from comparisons of multiple vertebrate genomes. *Nucleic Acids Res.*, **32**, W280–W286.
58. Poirot, O., Suhre, K., Abergel, C., O'Toole, E. and Notredame, C. (2004) 3DCoffee@igs: a web server for combining sequences and structures into a multiple sequence alignment. *Nucleic Acids Res.*, **32**, W37–W40.
59. Easton, D.F. (1994) Cancer risks in A-T heterozygotes. *Int. J. Radiat. Biol.*, **66**, S177–S182.
60. Thorstenson, Y.R., Roxas, A., Kroiss, R. et al. (2003) Contributions of ATM mutations to familial breast and ovarian cancer. *Cancer Res.*, **63**, 3325–3333.
61. Rodriguez-Lopez, R., Osorio, A., Sanchez-Pulido, L., De La Hoya, M., Barroso, A., Caldes, T. and Benitez, J. (2003) No mutations in the XRCC2 gene in BRCA1/2-negative high-risk breast cancer families. *Int. J. Cancer*, **103**, 136–137.
62. Smith, T.R., Miller, M.S., Lohman, K., Lange, E.M., Case, L.D., Mohrenweiser, H.W. and Hu, J.J. (2003) Polymorphisms of XRCC1 and XRCC3 genes and susceptibility to breast cancer. *Cancer Lett.*, **190**, 183–190.
63. Meijers-Heijboer, H., van den Ouweland, A., Klijn, J. et al. (2002) Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nat. Genet.*, **31**, 55–59.
64. Folias, A., Matkovic, M., Bruun, D., Reid, S., Hejna, J., Grompe, M., D'Andrea, A. and Moses, R. (2002) BRCA1 interacts directly with the Fanconi anemia protein FANCA. *Hum. Mol. Genet.*, **11**, 2591–2597.
65. Levran, O., Attwooll, C., Henry, R.T. et al. (2005) The BRCA1-interacting helicase BRIP1 is deficient in Fanconi anemia. *Nat. Genet.*, **37**, 931–933.
66. Venkitaraman, A.R. (2002) Connecting Fanconi's anaemia to breast cancer predisposition. *Lancet*, **360**, 1344–1345.
67. Collet-Durel, S., Guitton, N., Nourgalieva, K., Leveque, J., Danic, B. and Chenal, C. (2001) Genomic instability and breast cancer. *Oncol. Rep.*, **8**, 1001–1005.
68. Alter, B.P., Greene, M.H., Velazquez, I. and Rosenberg, P.S. (2003) Cancer in Fanconi anemia. *Blood*, **101**, 2072.
69. Van Der Heijden, M.S., Brody, J.R. and Kern, S.E. (2004) Functional screen of the fanconi anemia pathway in cancer cells by Fancd2 immunoblot. *Cancer Biol. Ther.*, **3**, 534–537.
70. Turner, N., Tutt, A. and Ashworth, A. (2004) Hallmarks of 'BRCAness' in sporadic cancers. *Nat. Rev. Cancer*, **4**, 814–819.
71. Lewis, A.G., Flanagan, J., Marsh, A., Pupo, G.M., Mann, G., Spurdle, A.B., Lindeman, G.J., Visvader, J.E., Brown, M.A. and Chenevix-Trench, G. (2005) Mutation analysis of FANCD2, BRIP1/BACH1, LMO4 and SFN in familial breast cancer. *Breast Cancer Res.*, **7**, R1005–R1016.
72. Holzel, M., van Diest, P.J., Bier, P., Wallisch, M., Hoatlin, M.E., Joenje, H. and de Winter, J.P. (2003) FANCD2 protein is expressed in proliferating cells of human tissues that are cancer-prone in Fanconi anaemia. *J. Pathol.*, **201**, 198–203.
73. Houghtaling, S., Timmers, C., Noll, M., Finegold, M.J., Jones, S.N., Meyn, M.S. and Grompe, M. (2003) Epithelial cancer in Fanconi anemia complementation group D2 (Fancd2) knockout mice. *Genes Dev.*, **17**, 2021–2035.
74. Houghtaling, S., Granville, L., Akkari, Y., Torimaru, Y., Olson, S., Finegold, M. and Grompe, M. (2005) Heterozygosity for p53 (Trp53+/–) accelerates epithelial tumor formation in fanconi anemia complementation group D2 (Fancd2) knockout mice. *Cancer Res.*, **65**, 85–91.
75. Seal, S., Barfoot, R., Jayatilake, H. et al. (2003) Evaluation of Fanconi Anemia genes in familial breast cancer predisposition. *Cancer Res.*, **63**, 8596–8599.
76. Thompson, E., Dragovic, R.L., Stephenson, S.A., Eccles, D.M., Campbell, I.G. and Dobrovic, A. (2005) A novel duplication polymorphism in the FANCA promoter and its association with breast and ovarian cancer. *BMC Cancer*, **5**, 43.
77. Gabriel, S.B., Schaffner, S.F., Nguyen, H. et al. (2002) The structure of haplotype blocks in the human genome. *Science*, **296**, 2225–2229.
78. Wang, N., Akey, J.M., Zhang, K., Chakraborty, R. and Jin, L. (2002) Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutation. *Am. J. Hum. Genet.*, **71**, 1227–1234.
79. Kim, W.Y. and Kaelin, W.G. (2004) Role of VHL gene mutation in human cancer. *J. Clin. Oncol.*, **22**, 4991–5004.
80. Otis, C.N., Krebs, P.A., Albuquerque, A., Quezada, M.M., San Juan, X., Sobel, M.E. and Merino, M.J. (2002) Loss of heterozygosity of p53, BRCA1, VHL, and estrogen receptor genes in breast carcinoma: correlation with related protein products and morphologic features. *Int. J. Surg. Pathol.*, **10**, 237–245.

Received November 29, 2005; revised March 21, 2006;
accepted April 19, 2006

Research article

Open Access

Genetic analysis of the vitamin D receptor gene in two epithelial cancers: melanoma and breast cancer case-control studies

Eva Barroso^{†1}, Lara P Fernandez^{†1}, Roger L Milne², Guillermo Pita³, Elena Sendagorta⁴, Uxua Floristan⁴, Marta Feito⁴, Jose A Aviles⁵, Manuel Martin-Gonzalez⁶, Jose I Arias⁷, Pilar Zamora⁷, Monserrat Blanco⁸, Pablo Lazaro⁵, Javier Benitez^{1,3} and Gloria Ribas^{*1}

Address: ¹Human Genetics Group; Human Cancer Genetics Program, Spanish National Cancer Research Centre (CNIO), Madrid, Spain, ²Genetic and Molecular Epidemiology Group; Human Cancer Genetics Program, CNIO, Madrid, Spain, ³National Genotyping Centre (CeGen), Human Cancer Genetics Program, CNIO, Madrid, Spain, ⁴Department of Dermatology, La Paz Hospital, Madrid, Spain, ⁵Department of Dermatology, Gregorio Marañon Hospital, Madrid, Spain, ⁶Department of Dermatology, Ramon y Cajal Hospital, Madrid, Spain, ⁷Service of Surgery, Monte Naranco, Oviedo, Spain and ⁸Department of Oncology, La Paz Hospital, Madrid, Spain

Email: Eva Barroso - eva.barroso@salud.madrid.org; Lara P Fernandez - lpfernandez@cnio.es; Roger L Milne - rlmilne@cnio.es; Guillermo Pita - gpita@cnio.es; Elena Sendagorta - elenasendagorta@hotmail.com; Uxua Floristan - uxuafloristan@hotmail.com; Marta Feito - marta8marta@hotmail.com; Jose A Aviles - jaavilesizquierdo@gmail.com; Manuel Martin-Gonzalez - martingolez@yahoo.es; Jose I Arias - joseignacio.arias@sespa.princast.es; Pilar Zamora - pzamora.hulp@salud.madrid.org; Monserrat Blanco - montseblancocodesido@yahoo.es; Pablo Lazaro - plazaroo@meditex.es; Javier Benitez - jbenitez@cnio.es; Gloria Ribas* - gribas@cnio.es

* Corresponding author †Equal contributors

Published: 23 December 2008

Received: 16 October 2008

BMC Cancer 2008, 8:385 doi:10.1186/1471-2407-8-385

Accepted: 23 December 2008

This article is available from: <http://www.biomedcentral.com/1471-2407/8/385>

© 2008 Barroso et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Vitamin D serum levels have been found to be related to sun exposure and diet, together with cell differentiation, growth control and consequently, cancer risk. Vitamin D receptor (VDR) genotypes may influence cancer risk; however, no epidemiological studies in sporadic breast cancer (BC) or malignant melanoma (MM) have been performed in a southern European population. In this study, the VDR gene has been evaluated in two epithelial cancers BC and MM.

Methods: We have conducted an analysis in 549 consecutive and non-related sporadic BC cases and 556 controls, all from the Spanish population, and 283 MM cases and 245 controls. Genotyping analyses were carried out on four putatively functional SNPs within the VDR gene.

Results: An association with the minor allele A of the non-synonymous SNP rs2228570 (rs10735810, *FokI*, Met1 Thr) was observed for BC, with an estimated odds ratio (OR) of 1.26 (95% CI = 1.02–1.57; $p = 0.036$). The synonymous variant rs731236 (*TaqI*) appeared to be associated with protection from BC (OR = 0.80, 95%CI = 0.64–0.99; $p = 0.047$). No statistically significant associations with MM were observed for any SNP. Nevertheless, sub-group analyses revealed an association between rs2228570 (*FokI*) and absence of childhood sunburns (OR = 0.65, $p = 0.003$), between the 3'utr SNP rs739837 (*BglII*) and fair skin (OR = 1.31, $p = 0.048$), and between the promoter SNP rs4516035 and the more aggressive tumour location in head-neck and trunk (OR = 1.54, $p = 0.020$).

Conclusion: In summary, we observed associations between SNPs in the VDR gene and BC risk, and a comprehensive analysis using clinical and tumour characteristics as outcome variables has revealed potential associations with MM. These associations required confirmation in independent studies.

BMC Cancer 2008, **8**:385

<http://www.biomedcentral.com/1471-2407/8/385>

Background

The vitamin D metabolite 1 α ,25-dihydroxivitamin D₃ (1,25D, also known as calcitriol) is the biologically active form of vitamin D₃ [1]. The concentration of vitamin D₃ in natural foods is quite low, and the majority of vitamin D₃ in individuals is from cholesterol metabolites in the skin upon exposure to ultraviolet (UV) radiation. 1,25D modulates the expression of specific genes in a tissue-specific manner by binding to the nuclear vitamin D receptor (VDR) and to specific DNA vitamin D response elements. The receptor and ligand induce a program of gene expression that contributes to the maintenance of the quiescent, differentiated phenotype. They are therefore able to regulate cellular proliferation, apoptosis and differentiation in many cell types [2].

Recent epidemiological studies have shown an association between low serum 1,25D levels and increased risk of breast, colorectal and prostate cancers. Furthermore, several studies have reported a possible link between polymorphic variants in the vitamin D receptor gene and increased susceptibility for primary and metastatic breast cancer, squamous cell carcinoma, colorectal cancer and prostate cancer [3-6]. Although the functional significance of these polymorphic variants remains unknown, there is strong evidence suggesting that they may have functional consequences in epithelial carcinogenesis and tumour progression [7,8]. VDR polymorphisms have been widely studied in Caucasian populations in relation to breast cancer (BC) [9-11] and malignant melanoma (MM) susceptibility [12,13], each finding different effects for SNPs, depending on the population analyzed and environmental factors acting upon them.

It is of general interest to study the most characterised variants in VDR in southern European countries, where sun exposure is typically higher than in northern European countries (maximum UV Index during the summer months = 9 in Spain versus 6.5 in Netherlands) [14,15]. In this study, we investigated for the first time the role of polymorphisms in VDR in two epithelial cancers, sporadic BC and MM, in the Spanish population. Additionally, clinical and tumour phenotypic variables have been taken into account to better define the involvement of VDR in these pathologies.

Methods

Study Subjects, Data Collection and DNA Extraction

BC Study

The BC case-control study included a total of 549 consecutive and non-related sporadic BC cases and 556 control women. Cases were recruited from 1st January 2002 to 31st December 2006 from three Spanish public hospitals: 258 (47%) from Monte Naranco Hospital, in Oviedo; 155 (28%) from the Fundación Jiménez Díaz, and 136 (25%) from La Paz University Hospital, both in Madrid. Con-

trols were unaffected Spanish women, recruited at three centres in Madrid: 455 (82%) from the Menopause Research Centre at the Instituto Palacios, 82 (15%) from the Fundación Jiménez Díaz, and 19 (3%) from the Madrid College of Lawyers. All cases and controls were women and controls were selected so that their age range was comparable to that of cases. We could not frequency match due to the larger numbers of cases.

Information about personal characteristics of cases and controls (age at diagnosis for cases or age at blood sample collection for controls, age at menarche, parity and menopausal status), and clinical and tumour characteristics for cases (metastasis at diagnosis, tumour grade, type and size, nodal involvement, and immunohistochemical markers), was either collected by the treating physician or extracted by review of medical records. This information is summarised in Additional file 1.

MM Study

The MM case-control study was based on 283 consecutive and non-related sporadic MM cases that were recruited from 1st September 2004 to 15th March 2008, at the Departments of Dermatology of three hospitals in Madrid: 147 (52%) from Gregorio Marañón General University Hospital, 54 (19%) from La Paz University Hospital and 82 (29%) from Ramón y Cajal University Hospital. A total of 245 cancer-free controls, frequency matched to cases by sex and age in ten-year categories, were recruited from the Madrid College of Lawyers (218 participants, 89%) and from Gregorio Marañón General University Hospital (27 participants, 11%).

A standardised questionnaire was used to collect information on pigmentation characteristics (eye colour, hair colour, skin colour, number of nevi, presence of solar lentigines), the presence of childhood sunburns, Fitzpatrick's classification of skin type, tumour location, Breslow index (deep index), and personal and family history of cancer, as described previously [16,17] (see Additional file 2). Fitzpatrick's classification of skin type was assessed for cases only, by review of medical records.

All participants in both studies were Caucasian and of Spanish origin. All subjects gave informed consent and the BC and MM studies were approved by the Ethics Committees of La Paz University Hospital and Gregorio Marañón General University Hospital, respectively.

Genomic DNA from cases and controls was extracted using the MagNA Pure LC Instrument according to the manufacturer's protocol as previously described [16,18,19].

SNP selection

Three public databases were used to collect information about SNPs in VDR: NCBI <http://www.ncbi.nlm.nih.gov>

Ensembl <http://www.ensembl.org>, and HapMap <http://www.hapmap.org>. Four SNPs were considered for inclusion because they have been widely analysed in previous epidemiological studies. All had minor allele frequency (MAF) greater than or equal to 10%. Two are located on exons, one is in the putative promoter region and the other in the 3'utr region. The two coding SNPs selected have been reported to be associated with breast cancer, in previous studies [11,20,21].

Genotyping assays

Genotyping was carried out using the TaqMan platform following the manufacturer's instructions. SNPs assays were designed using Applied Biosystems Assay-by-Design and Assay-on-Demand probes (Applied Biosystems, Foster City, CA, USA) (provided upon request). The genotype of each sample was automatically determined by measuring final allele-specific fluorescence in the ABI Prism 7900HT Detection System, using the SDS 2.1 software for allele discrimination (Applied Biosystems, Foster city, USA).

As a quality control measure, we included at least 2 sample duplicates and 1 non-template sample per 96-well plate. Genotypes were scored by two different personnel in the laboratory. We obtained a concordance rate of 100% for all four SNPs studied.

Statistical Analysis

For all polymorphisms studied, Fisher's exact test was used both to test for deviations from Hardy-Weinberg equilibrium (HWE) among controls and to compare differences in the MAF distributions between cases and controls.

In order to assess associations between genotypes, haplotypes and cancer risk, several analyses were performed. Genotype-related odds ratios (ORs), their corresponding 95% confidence intervals (CIs) and associated p-values were estimated via unconditional logistic regression. This was done for each of heterozygotes and minor-allele homozygotes relative to common-allele homozygotes, as well as under an additive model, in the latter case estimating an effect per copy of the minor allele carried. Known or suspected risk factors for BC (age, number of live births, age at menarche, and menopause status) and MM (eye colour, hair colour, skin colour, number of nevi, lentigines, and childhood sunburns) were evaluated for potential confounding effects by including them in multivariate analyses.

Associations between VDR polymorphisms genotyped and various individual, clinical and tumour characteristics were assessed via logistic regression in order to determine their potential modifying effects on BC and MM risk. This

was done for cases and controls pooled for each variable. Eye colour (blue/green versus brown), hair colour (blond/red versus brown/black), skin colour (fair versus brown), number of nevi (= 50 versus < 50), presence of lentigines (yes versus no) and childhood sunburn (yes versus no) were used as the outcome variables for MM.

Among BC cases only, the presence of metastatic disease at diagnosis (yes versus no), tumour histology (invasive versus in situ), tumour grade (grade > 1 versus grade 1), tumour size (> 2 cm versus = 2 cm), nodal involvement (yes versus no), estrogen receptor status (positive versus negative) and progesterone receptor status (positive versus negative), were used in the analysis. For MM cases-only analyses, the prior diagnosis of MM (yes versus no), phototype (I/II versus III/IV), tumour location (head/neck/trunk versus extremities) and tumour depth (T2/T3/T4 versus T0/T1) were considered as the outcome variables.

SPSS v11.0 was used to carry out these analyses. All p-values were two-sided and those less than 0.05 were considered statistically significant

Results and Discussion

Associations of VDR rs731236 and rs2228570 polymorphisms with cancer risk

Allelic frequencies for each SNP and the p-value for their comparison between cases and controls are presented in Table 1. We found no evidence of departure from Hardy-Weinberg equilibrium for any of the four SNPs genotyped (all p-values > 0.05). Results from univariate and multivariate genotype analysis are shown in Table 2.

We observed evidence of differences in minor allele frequency (MAF) between BC cases and controls for the synonymous change rs731236 (*TaqI*) ($p = 0.028$). The estimated OR per minor allele (C) in this SNP was 0.84 (95%CI 0.71–0.99, $p = 0.034$). This per-allele OR estimate was not substantially different in the multivariate analysis adjusting for age, number of live births, age at menarche, and menopause status (OR per allele = 0.85, 95% CI 0.69–1.03, $p = 0.102$). Regarding the SNP rs2228570 (*FokI*) (Met1Thr, formerly known as rs10735810), weak evidence of differences in MAF between BC cases and controls was observed ($p = 0.080$). The estimated OR per minor allele in this SNP was 1.17 (95%CI 0.98–1.40, $p = 0.081$), whereas the per-allele OR estimated in the multivariate analysis adjusting for potential confounding factors was higher, and statistically significant (OR per allele = 1.26, 95%CI 1.02–1.57, $p = 0.036$).

In general, previous studies have found no evidence of association with BC for rs731236 (*TaqI*) and rs2228570

BMC Cancer 2008, 8:385

<http://www.biomedcentral.com/1471-2407/8/385>

Table 1: Allelic frequencies comparison between cases and controls in the four SNPs tested, in both BC and MM pathologies

SNP ID	Other names	SNP Location	Nucleotide Change*	BREAST CANCER			MELANOMA			HapMap Caucasian MAF
				Cases (N = 549) MAF	Controls (N = 556) MAF	p-value**	Cases (N = 283) MAF	Controls (N = 245) MAF	p-value**	
rs4516035		5' upstream	T > C	0.41	0.39	0.27	0.44	0.40	0.28	0.45
rs2228570	<i>FokI</i>	MetI Thr	G > A	0.37	0.34	0.08	0.31	0.33	0.33	0.44
rs10735810										
rs731236	<i>TaqI</i>	Ile352	T > C	0.38	0.43	0.028	0.41	0.39	0.50	0.44
rs739837	<i>BglI</i>	3' utr	T > G	0.46	0.46	0.90	0.47	0.50	0.74	0.43

MAF, Minor Allele Frequency.

Statistically significant results ($p < 0.05$) indicated in bold.* Correspondence of nomenclature of SNP alleles are as following: the *FokI* alleles G and A correspond to F and f, respectively; the *TaqI* alleles T and C correspond to T and t, respectively; and *BglI* alleles T and G correspond to B and b, respectively.

**p-value, difference of MAF between cases and controls.

(*FokI*) [9,11,22-27]. All of these studies had limited statistical power to detect a moderate association. Population stratification may be another explanation for the lack of consistency in results. However, two studies with marginal statistically significant results for rs731236 (*TaqI*) reported contradictory results [11,24], whereas studies using larger sample sizes from Caucasian populations have shown risk effect of rs2228570 (*FokI*) consistent with that detected in the present study [20,21]. An association

with rs2228570 (*FokI*) was observed after adjustment for established risk factors including those used in the present study. This difference may be due to the tight relationship between VDR protein function and the hormonal aspect of BC aetiology such as menarche, parity and menopause [28,29].

In the case of MM, we did not observe any evidence of association with rs731236 (*TaqI*) (OR per allele = 1.08,

Table 2: Genotype frequencies comparison between cases and controls in the four SNPs tested, in both BC and MM pathologies

SNP ID	Statistical model	Genotype alleles	BREAST CANCER				MELANOMA			
			non-adjusted OR* (95% CI)	p-value	adjusted† OR* (95% CI)	p-value	non-adjusted OR* (95% CI)	p-value	adjusted‡ OR* (95% CI)	p-value
rs4516035	Codominant	CT	1.20 (0.92–1.56)	0.19	1.15 (0.83–1.58)	0.40	1.08 (0.73–1.60)	0.69	1.17 (0.70–1.96)	0.54
		CC	1.16 (0.82–1.65)	0.41	0.97 (0.63–1.49)	0.88	1.51 (0.90–2.53)	0.12	1.79 (0.91–3.53)	0.09
	Per minor allele	C-	1.09 (0.92–1.29)	0.30	1.01 (0.82–1.25)	0.91	1.20 (0.94–1.55)	0.15	1.31 (0.94–1.81)	0.11
rs2228570	Codominant	GA	1.14 (0.88–1.47)	0.32	1.26 (0.89–1.66)	0.22	1.09 (0.75–1.57)	0.66	1.23 (0.76–2.01)	0.40
		AA	1.41 (0.96–2.08)	0.08	1.65 (1.02–2.65)	0.041	0.69 (0.38–1.25)	0.22	1.23 (0.55–2.73)	0.61
	Per minor allele	A-	1.17 (0.98–1.40)	0.08	1.26 (1.02–1.57)	0.036	0.91 (0.70–1.19)	0.49	1.15 (0.81–1.64)	0.43
rs731236	Codominant	CT	0.84 (0.64–1.09)	0.19	0.82 (0.59–1.13)	0.22	1.26 (0.85–1.87)	0.25	1.06 (0.63–1.78)	0.81
		CC	0.70 (0.50–0.98)	0.040	0.72 (0.48–1.09)	0.13	1.09 (0.64–1.84)	0.76	1.17 (0.57–2.39)	0.66
	Per minor allele	C-	0.84 (0.71–0.99)	0.034	0.85 (0.69–1.03)	0.10	1.08 (0.84–1.40)	0.55	1.08 (0.77–1.52)	0.66
rs739837	Codominant	TG	0.96 (0.72–1.27)	0.76	0.96 (0.69–1.35)	0.83	0.89 (0.59–1.36)	0.60	0.64 (0.36–1.13)	0.12
		GG	1.03 (0.74–1.44)	0.86	1.30 (0.87–1.96)	0.20	0.82 (0.50–1.36)	0.45	0.69 (0.35–1.38)	0.29
	Per minor allele	G-	1.01 (0.86–1.20)	0.87	1.13 (0.92–1.38)	0.25	0.91 (0.71–1.17)	0.45	0.83 (0.59–1.16)	0.28

*OR: Odds Ratio estimated under codominant and log-additive models; CI: Confidence Interval.

†Adjusted for age at diagnosis, number of live births, age at menarche and menopause status.

‡Adjusted for eye colour, hair colour, skin colour, number of nevi, lentigines, and childhood sunburn.

Statistically significant results ($p < 0.05$) indicated in bold.

95%CI = 0.84–1.40, $p = 0.55$) or rs2228570 (*FokI*) (OR per allele = 0.91 95%CI = 0.70–1.19, $p = 0.49$). These results are consistent with the findings of smaller sample size studies [13,30,31] although other studies reported a protective tendency of rs731236 (*TaqI*) [12,32]. Only two studies reported an association of rs2228570 (*FokI*) and MM risk in North-European populations, one of them being a larger sample size study [12,32].

The rs731236 (*TaqI*) SNP is in linkage disequilibrium with other polymorphisms in the 3' extreme of the gene in Caucasian populations. Functional studies of these polymorphisms have evaluated their putative implication in the regulation of transcription, translation or RNA processing, but no consistent results were obtained [33]. However, functional studies of rs2228570 (*FokI*) have suggested a loss of the reported VDR benefits induced by the minor allele, due to its location in the first codon of the protein (Met1Thr). That is, the minor A allele appears to be associated with the use of an alternate start codon, which triggers a longer and less potent transcriptional activator protein form [7], which is consistent with it being associated with increased risk of BC. We did not observe evidence of association for any other SNP in BC or MM.

Associations of VDR polymorphisms with personal, clinical and tumoral characteristics

We assessed whether VDR SNPs were associated with various clinical and phenotypic characteristics using cases and controls combined. We tested for associations with tumour characteristics among cases only for each disease. Results are summarised in Table 3. The rare allele in the non-synonymous SNP rs2228570 (*TaqI*) appeared to be strongly associated with the absence of childhood sunburns (OR per allele = 0.65, 95% CI 0.49–0.86, $p = 0.003$), and this was maintained among controls only (OR per allele = 0.63). There was also weak evidence that it is associated with a prior diagnosis of MM in MM patients ($p = 0.060$). We also observed marginally significant associations for the 3'utr SNP rs739837 (*BglI*) with fair skin colour ($p = 0.048$) and with Fitzpatrick's photo-type I/II (0.070). Finally, the VDR promoter SNP rs4516035 was associated with tumours located in the head-neck or trunk ($p = 0.020$). No other associations were observed for BC or MM. Although we did not detect a significant effect of VDR SNPs directly on MM, the associations identified with MM phenotypic characteristics suggest that VDR SNPs may modulate MM susceptibility.

Further considerations

The strength of our study is the ability to control for the many established risk factors for BC and MM, although we recognize that there was potential for misclassification of phenotypic characteristics due to the subjective nature of

the phenotypic attributes considered. Controls participated on a volunteer basis which may have introduced some selection bias. However, the fact that they were frequency matched to cases on age and sex for melanoma and that breast cancer controls were selected so that their age range was comparable to that of cases and that the variable of primary interest was genetic would have kept such bias to a minimum. It should be noted that, the sample size of both studies was relatively limited and so associations can not be ruled out for rs731236 (*TaqI*), and rs2228570 (*FokI*), particularly in MM. The association of rs731236 (*TaqI*) and BC, was not statistically significant under a multivariate model. However, the estimated relative risk did not change substantially, indicating that the increase in p -value was due to the reduced sample size (with available covariate data), rather than due to confounding. Finally, conclusions are based on nominal p -values at 5% statistical significance and therefore require replication in independent studies.

Conclusion

BC and MM tumour pathologies may be influenced by the effect of variation in vitamin D intake through diet and sun exposure, together with VDR polymorphisms. Therefore, in a sunny region such as Spain, the effect of VDR polymorphisms on cancer risk may be more apparent than in other Caucasian populations. The results obtained in this study add to the evidence for a role of VDR as an important mediator in the development of cancer. We found evidence of association with BC for the non-synonymous variant rs2228570 (*FokI*), and for the synonymous SNP rs731236 (*TaqI*). These findings require replication in large samples and the role of these variants needs to be clarified by functional studies. We have also reported several associations among VDR SNPs and phenotypic risk factors that influence MM susceptibility, indicating their potential effect in disease development. The characterization of these and other polymorphisms in the VDR gene may help to better understand the aetiology and development of cancer, and to define risk groups to better target prevention strategies.

Abbreviations

VDR: Vitamin D receptor; BC: Breast cancer; MM: Malignant melanoma; SNP: Single nucleotide polymorphism; OR: Odds ratio; UV: Ultraviolet; DNA: Deoxyribonucleic acid; MAF: Minor allele frequency; HWE: Hardy-Weinberg equilibrium; CIs: Confidence Intervals

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

EB and LPF participated in the design of the study, acquisition of data, genotyping, analysis and interpretation of data, performed the statistical analysis and drafted the

BMC Cancer 2008, 8:385

<http://www.biomedcentral.com/1471-2407/8/385>**Table 3: Personal, clinical and tumoral phenotypic characteristics comparison from both BC and MM pathologies in the four SNPs tested**

Tumour Type	Characteristic	rs4516035 OR* (95% CI)	p-value	rs2228570 OR* (95% CI)	p-value	rs731236 OR* (95% CI)	p-value	rs739837 OR* (95% CI)	p-value
BREAST CANCER									
	Metastasis†	0.53 (0.25–1.11)	0.09	1.00 (0.51–1.96)	0.99	1.20 (0.63–2.27)	0.58	1.11 (0.57–2.17)	0.75
	Tumor histology (Invasive)†	0.87 (0.57–1.32)	0.51	0.83 (0.54–1.28)	0.40	1.10 (0.72–1.69)	0.66	0.75 (0.49–1.15)	0.19
	Tumor grade (Grade > I)†	0.93 (0.67–1.30)	0.68	1.19 (0.84–1.68)	0.33	1.22 (0.87–1.70)	0.26	0.76 (0.54–1.05)	0.09
	Tumor size (> 2 cm)†	1.02 (0.77–1.35)	0.92	0.98 (0.73–1.30)	0.88	1.14 (0.86–1.51)	0.37	0.81 (0.61–1.08)	0.15
	Nodal involvement†	0.93 (0.70–1.23)	0.61	0.95 (0.71–1.27)	0.72	0.94 (0.71–1.25)	0.69	1.09 (0.82–1.45)	0.54
	ER positive†	0.96 (0.66–1.41)	0.85	1.24 (0.84–1.84)	0.28	1.31 (0.90–1.90)	0.16	1.02 (0.71–1.46)	0.91
	PR positive†	0.92 (0.67–1.26)	0.60	1.12 (0.80–1.55)	0.51	0.91 (0.67–1.24)	0.56	1.12 (0.83–1.51)	0.48
MELANOMA									
	Light Eye Colour+	1.03 (0.79–1.34)	0.82	0.91 (0.69–1.21)	0.52	0.92 (0.70–1.21)	0.56	1.06 (0.81–1.39)	0.64
	Blond/Red Hair Colour+	0.99 (0.71–1.40)	0.99	0.78 (0.55–1.13)	0.19	1.02 (0.73–1.44)	0.89	1.45 (0.89–1.74)	0.20
	Fair Skin Colour+	0.97 (0.75–1.26)	0.83	1.08 (0.82–1.42)	0.57	0.86 (0.66–1.12)	0.25	1.31 (1.00–1.70)	0.048
	N° of Nevi = 50+	0.97 (0.66–1.40)	0.85	1.00 (0.67–1.49)	0.99	1.06 (0.72–1.54)	0.77	1.34 (0.90–1.97)	0.15
	Presence of Lentigines+	1.03 (0.79–1.34)	0.84	0.79 (0.59–1.05)	0.11	1.20 (0.91–1.58)	0.21	0.90 (0.69–1.18)	0.45
	Presence of Childhood Sunburns+	1.12 (0.87–1.45)	0.39	0.65 (0.49–0.86)	0.003	1.10 (0.84–1.43)	0.49	0.91 (0.70–1.19)	0.50
	Other MM†	1.26 (0.52–3.03)	0.61	2.43 (0.95–6.19)	0.060	1.83 (0.73–4.56)	0.19	0.70 (0.26–1.85)	0.47
	Fitzpatrick's phototype I/II†	0.74 (0.51–1.09)	0.13	0.85 (0.63–1.46)	0.89	0.91 (0.62–1.34)	0.63	1.43 (0.97–2.12)	0.070
	Tumor Location (Head/Neck/Trunk)†	1.54 (1.08–2.20)	0.020	1.16 (0.79–1.72)	0.45	1.16 (0.81–1.65)	0.42	1.17 (0.82–1.67)	0.38
	Breslow Index (T2/T3/T4)†	1.04 (0.72–1.51)	0.82	1.20 (0.80–1.79)	0.38	0.84 (0.57–1.24)	0.38	1.01 (0.69–1.47)	0.96

*OR: Odds Ratio per minor allele; CI: Confidence Interval, unadjusted p-values.

+Cases and controls pooled for each variable.

†Cases only considered.

Statistically significant results ($p < 0.05$) indicated in bold.

manuscript. RLM performed the interpretation of data, the statistical analysis and drafted the manuscript. GP has been involved in the acquisition, designing sample database used for the study and interpretation of data. ES, UF, MF, JAA, MM, JIA, PZ, MB and PL participated in patient enrolment and patient phenotypic classification and the acquisition of clinical data. JB have been involved in revising the manuscript and in given final approval of the version to be published. GR participated in the design of the study, analysis and interpretation of data, revising the manuscript and in given final approval of the version to be published.

The manuscript has been seen and approved by all listed authors. E Barroso and LP Fernandez contributed equally to this work.

Additional material

Additional file 1

Personal, clinical and tumoral phenotypic characteristics in cases and controls in BC. The data provided represent the personal, clinical and tumoral phenotypic characterization of BC samples used in the study.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2407-8-385-S1.doc>]

Additional file 2

Personal, clinical and tumoral phenotypic characteristics in cases and controls in MM. The data provided represent the personal, clinical and tumoral phenotypic characterization of MM samples used in the study. Click here for file
[<http://www.biomedcentral.com/content/supplementary/1471-2407-8-385-S2.doc>]

Acknowledgements

This study was supported by grants BFI2003-03852 and SAF2007-65542-C02-01 from the Ministerio de Educación y Ciencia (MEC) and Fundación Mútua Madrileña, Spain (GR). EB is funded by the Comunidad Autónoma de Madrid and LPF is funded by the Ministerio de Sanidad y Consumo under a grant from the Fondo de Investigación Sanitaria FI05/00918. We would like to thank Álvaro Ruibal (University Hospital, Santiago de Compostela), and Santiago Palacios (Instituto Palacios, Madrid), Mariano Casado, Angel Pizarro and Matias Mayor (Hospital La Paz), Angeles de la Riva Grandal (Hospital Ramón y Cajal) and staff at the Hospital Gregorio Marañón and Madrid College of Lawyers for the access to samples of cases and controls. We would also like to thank Fátima Mercadillo, Alicia Barroso, Victoria Fernández and Rocio Letón for their expert technical skills.

References

- Trang HM, Cole DE, Rubin LA, Pierratos A, Siu S, Vieth R: Evidence that vitamin D3 increases serum 25-hydroxyvitamin D more efficiently than does vitamin D2. *Am J Clin Nutr* 1998, **68**(4):854-858.
- Jurutka PW, Whitfield GK, Hsieh JC, Thompson PD, Haussler CA, Haussler MR: Molecular nature of the vitamin D receptor and its role in regulation of gene expression. *Rev Endocr Metab Disord* 2001, **2**(2):203-216.
- Deeb KK, Trump DL, Johnson CS: Vitamin D signalling pathways in cancer: potential for anticancer therapeutics. *Nat Rev Cancer* 2007, **7**(9):684-700.
- Bikle DD, Oda Y, Xie Z: Vitamin D and skin cancer: a problem in gene regulation. *J Steroid Biochem Mol Biol* 2005, **97**(1-2):83-91.
- Cui Y, Rohan TE: Vitamin D, calcium, and breast cancer risk: a review. *Cancer Epidemiol Biomarkers Prev* 2006, **15**(8):1427-1437.
- Guy M, Lowe LC, Bretherton-Watt D, Mansi JL, Peckitt C, Bliss J, Wilson RG, Thomas V, Colston KW: Vitamin D receptor gene polymorphisms and breast cancer risk. *Clin Cancer Res* 2004, **10**(16):5472-5481.
- Arai H, Miyamoto K, Taketani Y, Yamamoto H, Iemori Y, Morita K, Tonai T, Nishisho T, Mori S, Takeda E: A vitamin D receptor gene polymorphism in the translation initiation codon: effect on protein activity and relation to bone mineral density in Japanese women. *J Bone Miner Res* 1997, **12**(6):915-921.
- Colin EM, Weel AE, Uitterlinden AG, Buurman CJ, Birkenhager JC, Pols HA, van Leeuwen JP: Consequences of vitamin D receptor gene polymorphisms for growth inhibition of cultured human peripheral blood mononuclear cells by 1, 25-dihydroxyvitamin D3. *Clin Endocrinol (Oxf)* 2000, **52**(2):211-216.
- McCullough ML, Stevens VL, Diver WR, Feigelson HS, Rodriguez C, Bostick RM, Thun MJ, Calle EE: Vitamin D pathway gene polymorphisms, diet, and risk of postmenopausal breast cancer: a nested case-control study. *Breast Cancer Res* 2007, **9**(1):R9.
- Trabert B, Malone KE, Daling JR, Doody DR, Bernstein L, Ursin G, Marchbanks PA, Strom BL, Humphrey MC, Ostrander EA: Vitamin D receptor polymorphisms and breast cancer risk in a large population-based case-control study of Caucasian and African-American women. *Breast Cancer Res* 2007, **9**(6):R84.
- Abbas S, Nieters A, Linseisen J, Slinger T, Kropp S, Mutschelknauss EJ, Flesch-Janys D, Chang-Claude J: Vitamin D receptor gene polymorphisms and haplotypes and postmenopausal breast cancer risk. *Breast Cancer Res* 2008, **10**(2):R31.
- Hutchinson PE, Osborne JE, Lear JT, Smith AG, Bowers PW, Morris PN, Jones PW, York C, Strange RC, Fryer AA: Vitamin D receptor polymorphisms are associated with altered prognosis in patients with malignant melanoma. *Clin Cancer Res* 2000, **6**(2):498-504.
- Santonico C, Capizzi R, Concolino P, Lavieri MM, Paradisi A, Gentileschi S, Torti E, Rutella S, Rocchetti S, Di Carlo A, et al.: Association between cutaneous melanoma, Breslow thickness and vitamin D receptor Bsm1 polymorphism. *Br J Dermatol* 2007, **156**(2):277-282.
- Den Outer PN, Slaper H, Tax RB: UV radiation in the Netherlands: Assessing long-term variability and trends in relation to ozone and clouds. *Journal of geophysical research* 2005, **110**:D02203. DOI:10.1029/2004JD005111
- Marin MJ, Sola Y, Tena F, Utrillas MP, Campmany E, de Cabo X, Lorente J, Martinez-Lozano JA: The UV index on the Spanish Mediterranean coast. *Photochemistry and photobiology* 2005, **81**(3):659-665.
- Fernandez L, Milne R, Bravo J, Lopez J, Aviles J, Longo M, Benitez J, Lazaro P, Ribas G: MC1R: three novel variants identified in a malignant melanoma association study in the Spanish population. *Carcinogenesis* 2007, **28**(8):1659-1664.
- Fernandez LP, Milne RL, Pita G, Aviles JA, Lazaro P, Benitez J, Ribas G: SLC45A2: a novel malignant melanoma-associated gene. *Hum Mutat* 2008, **29**(9):1161-1167.
- Kessler HH, Muhlbauer G, Stelzl E, Daghofer E, Santner BI, Marth E: Fully automated nucleic acid extraction: MagNA Pure LC. *Clin Chem* 2001, **47**(6):1124-1126.
- Barroso E, Milne RL, Fernandez LP, Zamora P, Arias JL, Benitez J, Ribas G: FANCD2 associated with sporadic breast cancer risk. *Carcinogenesis* 2006, **27**(9):1930-1937.
- Chen WY, Bertone-Johnson ER, Hunter DJ, Willett WC, Hankinson SE: Associations between polymorphisms in the vitamin D receptor and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 2005, **14**(10):2335-2339.
- Gapska P, Scott RJ, Serrano-Fernandez P, Huzarski T, Byrski T, Kladny J, Gronwald J, Gorski B, Cybulski C, Lubinski J, et al.: Vitamin D receptor variants and breast cancer risk in the Polish population. *Breast Cancer Res Treat* 2008.
- Hou MF, Tien YC, Lin GT, Chen CJ, Liu CS, Lin SY, Huang TJ: Association of vitamin D receptor gene polymorphism with sporadic breast cancer in Taiwanese patients. *Breast Cancer Res Treat* 2002, **74**(1):1-7.
- Buyru N, Tezol A, Yosunkaya-Fenerci E, Dalay N: Vitamin D receptor gene polymorphisms in breast cancer. *Exp Mol Med* 2003, **35**(6):550-555.
- Curran JE, Vaughan T, Lea RA, Weinstein SR, Morrison NA, Griffiths LR: Association of A vitamin D receptor polymorphism with sporadic breast cancer development. *Int J Cancer* 1999, **83**(6):723-726.
- Bretherton-Watt D, Given-Wilson R, Mansi JL, Thomas V, Carter N, Colston KW: Vitamin D receptor gene polymorphisms are associated with breast cancer risk in a UK Caucasian population. *Br J Cancer* 2001, **85**(2):171-175.
- Ingles SA, Garcia DG, Wang W, Nieters A, Henderson BE, Kolonel LN, Haile RW, Coetzee GA: Vitamin D receptor genotype and breast cancer in Latinas (United States). *Cancer Causes Control* 2000, **11**(1):25-30.
- John EM, Schwartz GG, Koo J, Wang W, Ingles SA: Sun exposure, vitamin D receptor gene polymorphisms, and breast cancer risk in a multiethnic population. *Am J Epidemiol* 2007, **166**(12):1409-1419.
- Gilad LA, Bresler T, Gnainsky J, Smirnoff P, Schwartz B: Regulation of vitamin D receptor expression via estrogen-induced activation of the ERK 1/2 signaling pathway in colon and breast cancer cells. *J Endocrinol* 2005, **185**(3):577-592.
- Gilad LA, Schwartz B: Association of estrogen receptor beta with plasma-membrane caveola components: implication in control of vitamin D receptor. *J Mol Endocrinol* 2007, **38**(6):603-618.
- Halsall JA, Osborne JE, Potter L, Pringle JH, Hutchinson PE: A novel polymorphism in the 1A promoter region of the vitamin D receptor is associated with altered susceptibility and prognosis in malignant melanoma. *Br J Cancer* 2004, **91**(4):765-770.
- Han J, GA Colditz, DJ Hunter: Polymorphisms in the MTHFR and VDR genes and skin cancer risk. *Carcinogenesis* 2007, **28**:390-7.

BMC Cancer 2008, **8**:385

<http://www.biomedcentral.com/1471-2407/8/385>

32. Li C, Z Liu, LE Wang, JE Gershenwald, JE Lee, VG Prieto, *et al.*: **Haplotype and genotypes of the VDR gene and cutaneous melanoma risk in non-Hispanic whites in Texas: a case-control study.** *International Journal of Cancer* 2008, **122**:2077-84.
33. Durrin LK, Haile RW, Ingles SA, Coetzee GA: **Vitamin D receptor 3'-untranslated region polymorphisms: lack of effect on mRNA stability.** *Biochim Biophys Acta* 1999, **1453**(3):311-320.

Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-2407/8/385/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

